Λ-MINIMAX ESTIMATION I: TWO TERMINAL ACTION PROBLEMS

BU-278-M Daniel L. Solomon June, 1970

Abstract

Statistical decision problems are considered in which the decision maker is assumed to have prior information but cannot completely specify a prior distribution. The decision maker's prior knowledge is reflected in his willingness to specify a subset, Λ of the class of all prior distributions, Λ^*. He is then recommended to select a decision rule to minimize the maximum over distributions in Λ of the Bayes risk. Such a rule is called Λ-minimax and reduces to a Bayes rule with respect to λ if $\Lambda = \{\lambda\}$ and a minimax rule if $\Lambda = \Lambda^*$. This paper is concerned with Λ-minimax estimation of a location parameter from an arbitrary multivariate distribution and with estimation of the variance of a (univariate) Normal distribution.

Biometrics Unit Mimeo Series, Cornell University, Ithaca, New York 14850.

Λ-MINIMAX ESTIMATION I: TWO TERMINAL ACTION PROBLEMS

Daniel L. Solomon

Statistical decision problems are considered in which the decision maker is assumed to have prior information but cannot completely specify a prior distribution. The decision maker's prior knowledge is reflected in his willingness to specify a subset, Λ of the class of all prior distributions, Λ^*. He is then recommended to select a decision rule to minimize the maximum over distributions in Λ of the Bayes risk. Such a rule is called Λ-minimax and reduces to a Bayes rule with respect to λ if $\Lambda = \{\lambda\}$ and a minimax rule if $\Lambda = \Lambda^*$. This paper is concerned with Λ-minimax estimation of a location parameter from an arbitrary multivariate distribution and with estimation of the variance of a (univariate) Normal distribution.

1. INTRODUCTION

In statistical decision problems, the decision maker may wish to choose between Bayes and minimax decision procedures. Even if he prefers Bayesian techniques, he is often confronted with an incomplete specification of his prior distribution and thus cannot apply them. This paper suggests an alternative which is other than minimax and recommends that the decision maker apply Bayes criteria as far as his prior knowledge permits. This will be made precise and some particular problems considered.

In the case of incomplete prior information, the decision maker can often improve his a priori knowledge (perhaps at some cost or utility loss) by, for example, introspection, consultation or an interviewing technique such as that due to Winkler [17] for eliciting a prior distribution. It will be proposed in a sequel to this paper that this possibility be included in the cost structure of the model. To this author's knowledge, this aspect of the problem has not been investigated. Several authors have, however, considered the terminal action problem in the presence of incomplete prior information.

Suppose that the decision maker's prior information takes the form of a best guess θ_o of a parameter θ which is to be estimated. Thompson [15] specifies a method for "shrinking" the usual estimator $\hat{\theta}$ (maximum likelihood, minimum variance unbiased, etc.) toward θ_o by using an estimator of the form

$$\hat{\theta}_s = c\hat{\theta} + (1 - c)\theta_o .$$

Stone [15] suggests restricting the class of decision rules in such a way that only approximate prior information is required. Approaches to combining Bayes and minimax procedures are given by Hodges and Lehmann [6], Skibinsky and Cote [13], Schneeweiss [12], Blum and Rosenblatt [2], Randles [10], George [3], and others.

Now consider a criterion for combining Bayes and minimax procedures due to Menges [8]. The criterion is adopted for the analysis of this paper. For a given decision maker, assume he has a prior distribution, λ, but complete specification of λ is unavailable to him. Assume that he does however have some prior information reflected in his willingness to assert that $\lambda \epsilon \Lambda \subset \Lambda^*$, where Λ^* is the class of all probability measures on the parameter space Ω.

Menges recommends that the decision maker is to choose a rule δ_o such that

$$\sup_{\lambda \epsilon \Lambda} r^*(\lambda, \delta_o) = \inf_{\delta} \sup_{\lambda \epsilon \Lambda} r^*(\lambda, \delta), \tag{1}$$

where $r^*(\lambda, \delta)$ is the Bayes risk in δ with respect to the prior distribution λ. He calls such a rule δ_o, an extended Bayes rule, here called a Λ-minimax rule after

Blum and Rosenblatt [2].

The Bayesian character of the criterion is in the minimization of the Bayes risk r^*, while its minimax nature is in choosing δ_0 so that the maximum Bayes risk is as small as possible regardless of the prior distribution $\lambda \epsilon \Lambda$. Should Λ contain only one distribution λ, then a Λ-minimax rule is a Bayes rule with respect to that λ. On the other hand, if $\Lambda = \Lambda^*$, the class of all prior distributions, then the maximum is attained at that λ which assigns probability one to the values of θ which maximize the risk, r. Thus

$$\sup_{\lambda \epsilon \Lambda^*} r^*(\lambda,\delta) = \sup_{\theta \epsilon \Omega} r(\theta,\delta),$$

and a Λ^*-minimax rule is a minimax rule.

This paper applies Menges' approach to the particular problems of estimation of a multivariate location parameter and estimation of a normal scale parameter. A sequel considers some problems concerning optimal designs.

To motivate the discussion, consider the following example. Suppose a value x is observed of a random variable X, which has mean $\theta \epsilon \Omega = (-\infty,\infty)$ and known variance σ^2, and a linear estimator δ of θ is sought. That is, the class of decision rules of interest is restricted to

$$D_L = \{\delta \epsilon D | \delta(x) = bx + c\} ,$$

where D is the class of all non-randomized decision rules. Suppose further that the loss associated with this decision is $(\delta(x) - \theta)^2$ if θ is the state of nature. In the absence of other information the minimax criterion might be applied; that is choose b_1, c_1 such that

$$\sup_{\theta \epsilon \Omega} E\{(b_1 X + c_1 - \theta)^2 | \theta\} = \inf_{b,c} \sup_{\theta \epsilon \Omega} E\{(bX + c - \theta)^2 | \theta\} .$$

It can be shown that $b_1 = 1$, $c_1 = 0$ so that $\delta_1(x) = x$ is minimax among the class of linear rules, D_L.

Consider now a situation in which the information is available that θ is in some bounded interval, say $|\theta - \Delta| \leq M$, where Δ and M are known, $M \geq 0$. How should this information be used to modify the estimate $\delta(x)$? With the minimax criterion the maximization is restricted to the set $\{\theta \epsilon \Omega \,\big|\, |\theta - \Delta| \leq M\}$. For $\delta \epsilon D_L$ one can show that the minimax rule is now

$$\delta_2(x) = b_2 x + c_2$$

where

$$b_2 = \frac{M^2}{\sigma^2 + M^2}, \quad c_2 = \frac{\sigma^2 \Delta}{\sigma^2 + M^2} = (1 - b_2)\Delta \; . \tag{2}$$

Now the risk in the minimax rule $\delta_1(x) = x$ is

$$r(\theta, \delta_1) = E\{(X - \theta)^2 | \theta\} = \sigma^2 \; ,$$

while from (2) and the fact that $|\theta - \Delta| \leq M$

$$\begin{aligned} r(\theta, \delta_2) &= E\{(b_2 X + c_2 - \theta)^2 | \theta\} \\ &= \left(\frac{M^2}{\sigma^2 + M^2} \right)^2 \sigma^2 + \left(\frac{\sigma^2}{\sigma^2 + M^2} \right)^2 (\Delta - \theta)^2 \\ &\leq \frac{M^2 \sigma^2}{\sigma^2 + M^2} \leq \sigma^2 = r(\theta, \delta_1) \; , \end{aligned}$$

with strict inequality if $M < \infty$ and $\sigma^2 \neq 0$. Thus if it is known that $|\theta - \Delta| \leq M$, the risk in δ_2 is no larger than that in δ_1, and is smaller in most situations.

Assume now, that a prior distribution, λ, is available for θ of the previous example with

$$E_\lambda \theta = \mu \quad \text{and} \quad \text{var}_\lambda \theta = \sigma_o^2 \; .$$

The linear Bayes estimate of θ with respect to λ is

$$\delta_o(x) = \frac{\frac{x}{\sigma^2} + \frac{\mu}{\sigma_o^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_o^2}} \; ,$$

with Bayes risk

$$\left[\frac{1}{\sigma^2} + \frac{1}{\sigma_o^2} \right]^{-1} .$$

On the other hand, suppose that the prior information is incomplete, and the decision maker knows σ_o^2 but only that $|\Delta - \mu| \leq M$. Then it is reasonable to choose a linear decision rule δ_3 such that

$$\sup_{\lambda \in \Lambda} E_\lambda \{ E[(\delta_3(X) - \theta)^2 | \theta] \} = \inf_{\delta \in D_L} \sup_{\lambda \in \Lambda} E_\lambda \{ E[(\delta(X) - \theta)^2 | \theta] \} ,$$

where Λ is the set of all prior distributions for which $|\Delta - \mu| \leq M$. It will be shown later in this paper that

$$\delta_3(x) = \frac{\dfrac{x}{\sigma^2} + \dfrac{\Delta}{\sigma_o^2 + M^2}}{\dfrac{1}{\sigma^2} + \dfrac{1}{\sigma_o^2 + M^2}}$$

and that the risk of δ_3 is

$$\sup_{\lambda \in \Lambda} E_\lambda \{ E[(\delta_3(X) - \theta)^2 | \theta] \} = \left[\frac{1}{\sigma^2} + \frac{1}{\sigma_o^2 + M^2} \right]^{-1} \leq \sigma^2 .$$

Defining

$$r_0 = E_\lambda r(\theta, \delta_o) = \left[\frac{1}{\sigma^2} + \frac{1}{\sigma_o^2} \right]^{-1} ,$$

$$r_1 = r(\theta, \delta_1) = \sigma^2 ,$$

$$r_2 = \underset{|\theta - \Delta| \leq M}{\text{supremum}} \, r(\theta, \delta_2) = \left[\frac{1}{\sigma^2} + \frac{1}{M^2} \right]^{-1} ,$$

$$r_3 = \underset{\lambda \in \Lambda}{\text{supremum}} \, E_\lambda r(\theta, \delta_3) = \left[\frac{1}{\sigma^2} + \frac{1}{\sigma_o^2 + M^2} \right]^{-1} ,$$

it can be shown that

$$r_0 \leq r_2 \leq r_3 \leq r_1 \quad \text{if} \quad \sigma_o^2 \leq M^2$$

while

$$r_2 \leq r_0 \leq r_3 \leq r_1 \quad \text{if} \quad \sigma_o^2 \geq M^2 .$$

2. Λ-MINIMAX ESTIMATES OF LOCATION

2.1 The Model

Suppose that X and $\bar{\underset{-}{\theta}}$ are real (p × 1) vector random variables (not necessarily Normal), that $\mu' = (\mu_1', \mu_2', \cdots, \mu_p')^T$ and $\Delta = (\Delta_1, \Delta_2, \cdots, \Delta_p)^T$ are (p × 1) vectors of real constants, $M = (m_1, m_2, \cdots, m_p)^T$ is a (p × 1) vector of non-negative real constants, and that $\Sigma_0 = (\sigma_{ij}^o)$, $\Sigma_1 = (\sigma_{ij}^1)$ and $K = (k_{ij})$ are (p × p) symmetric positive definite matrices of real constants. Assume further that the distribution of X is such that

$$E(X|\bar{\underset{-}{\theta}}) = \bar{\underset{-}{\theta}} \quad \text{and} \quad \mathrm{Cov}(X|\bar{\underset{-}{\theta}}) = \Sigma_1$$

where Σ_1 is known.

Let D be the class of (decision) functions from E^p (p-dimensional Euclidean space) into E^p. Then for $\delta \epsilon D$, $\delta(X)$ is an "estimator" of the location "parameter" $\bar{\underset{-}{\theta}}$. Suppose that if X is observed to have the value x, and $\bar{\underset{-}{\theta}}$ is estimated by $\delta(x)$, when $\bar{\underset{-}{\theta}} = \theta$, then the loss incurred is

$$\ell(\delta, x, \theta) \equiv (\delta(x) - \theta)^T K(\delta(x) - \theta) .$$

Suppose that the decision maker's prior knowledge is such that his prior distribution λ on Ω has

$$E_{\lambda}\bar{\underset{-}{\theta}} = \mu' \quad \text{and} \quad \mathrm{Cov}_{\lambda}\bar{\underset{-}{\theta}} = \Sigma_0 , \tag{3}$$

where Σ_0 is known, and that he has learned that

$$\mu' \epsilon U' = \{\mu' \epsilon E^p \,\big|\, |\Delta - \mu'| \leq M\} , \tag{4}$$

where for p-vectors Y and Z, $|Y| \leq Z$ means $|y_i| \leq z_i$, $i = 1, 2, \cdots, p$. Then take for Λ, that class of prior distributions which satisfy (3) with $\mu' \epsilon U'$. It is noted at the end of section 2.4 that the results of this section apply for a more general formulation of U'. In particular, the results hold whenever U' is the convex hull of a finite number of points. We choose to work with the "box" for simplicity of presentation.

We make some definitions and find an optimality criterion for the estimator.

For $\delta \epsilon D^* \subset D$, define the <u>Bayes risk of δ for given μ'εU'</u> by

$$\begin{aligned} B(\mu',\delta) &= B(\mu',\delta,\Delta,M) \equiv E\{E[\ell(\delta,X,\underset{\sim}{\bar{\theta}})|\underset{\sim}{\bar{\theta}}]\} \\ &= E\{E[(\delta(X) - \underset{\sim}{\bar{\theta}})^T K(\delta(X) - \underset{\sim}{\bar{\theta}})|\underset{\sim}{\bar{\theta}}]\} \end{aligned}, \tag{5}$$

the <u>maximum Bayes risk of δ</u> by

$$B(\delta) = \sup_{\mu'\epsilon U'} B(\mu',\delta) ,$$

and the <u>Λ-minimax risk</u> by

$$B^*(U') = \inf_{\delta\epsilon D^*} B(\delta) .$$

Any rule $\delta_o \epsilon D^*$ for which

$$B(\delta_o) = B^*(U') ,$$

or equivalently for which

$$\sup_{\mu'\epsilon U'} B(\mu',\delta_o) = \inf_{\delta\epsilon D^*} \sup_{\mu'\epsilon U'} B(\mu',\delta) ,$$

is then said to be <u>Λ-minimax in D*</u>. Interchangeably, δ_o will be termed a U'-minimax rule.

Thus for each $\delta \epsilon D^*$, determine a worst (in terms of Bayes risk) μ' and choose a δ_o for which the maximum risk is a minimum. This paper is concerned with finding a Λ-minimax rule in D_L, defined by

$$D_L \equiv \{\delta \in D \mid \delta(x) = Bx + C\}$$

where $B(p \times p)$ and $C(p \times 1)$ are matrices of real constants which may depend on Δ and M. That is, D_L is the class of linear functions from E^p into E^p. Summarizing, a $\underline{\Lambda}$- (or U'-) <u>minimax rule in</u> $\underline{D_L}$ is a rule

$$\delta_o(x) = B_o x + C_o$$

such that

$$\sup_{\mu' \epsilon U'} B(\mu', \delta_o) = \inf_{\delta \epsilon D_L} \sup_{\mu' \epsilon U'} B(\mu', \delta) .$$

This criterion is related to the usual minimax criterion as follows:

$$\begin{aligned} B^*(U') &= \inf_{\delta \epsilon D^*} \sup_{\mu' \epsilon U'} E\{E[(\delta(X) - \underline{\bar{\theta}})^T K(\delta(X) - \underline{\bar{\theta}}) \mid \underline{\bar{\theta}}]\} \\ &= \inf_{\delta \epsilon D^*} \sup_{\mu' \epsilon U'} E\{E[(\delta(X) - \underline{\bar{\theta}})^T K(\delta(X) - \underline{\bar{\theta}}) \mid X, \mu']\} \\ &= \inf_{\delta \epsilon D^*} \sup_{\mu' \epsilon U'} E\, C(\mu', \delta(X)) , \end{aligned}$$

where

$$C(\mu', \delta(X)) = E[(\delta(X) - \underline{\bar{\theta}})^T K(\delta(X) - \underline{\bar{\theta}}) \mid X, \mu'] .$$

This is the minimax criterion for loss function C and state space U'.

Consider the following hypothetical applications of the problem just described.

The National Bureau of Standards maintains the standard for the unit of electromotive force (emf), the volt. This standard is determined from the average value, μ', of 44 saturated Weston cells which comprise the National Reference Group (NRG).

Occasionally prototypes of the NRG cells are made with emf value $\theta \sim N(\mu', \sigma_o^2)$. These cells are used in turn as models for customers' cells with values $X \sim N(\theta, \sigma_1^2)$.

Over the years, values of σ_0^2 and σ_1^2 have been established.

It has been observed that μ' varies in time. By costly physical experiments, bounds can be placed on μ', say $\mu' \epsilon U'$, an interval. A customer wishes to estimate the θ for his prototype. His class of prior distributions for θ is then $N(\mu', \sigma_0^2)$, $\mu' \epsilon U'$. For more detail on these techniques see Hamer [5, pp. 6-7].

As a second application and in a different context, suppose that observations of a particular individual's serum cholesterol at a given time are $X \sim N(\theta, \sigma_1^2)$. The parameter $\theta \epsilon \Omega$ varies from person to person and suppose that $\underset{\sim}{\bar{\theta}} \sim N(\mu', \sigma_0^2)$, where μ' is the mean cholesterol level for the population at large. Now at various costs, an individual's age, height, weight, diet, medical history, etc., can be determined and thus the individual can be placed in a sub-population. If it is now supposed that for this sub-population it is known that the mean cholesterol level μ' is in some set U', then to estimate θ, the class of prior distributions is $N(\mu', \sigma_0^2)$, $\mu' \epsilon U'$.

2.2 Existence

Theorem 9 of section 2.4 asserts that in the class D_L of linear decision rules, any Λ-minimax rule is in $L \subset D_L$ defined by

$$L = \{\delta \epsilon D_L | \delta(x) = Bx + \bar{B}\Delta\} , \tag{6}$$

where

$$\bar{B} = (I - B) .$$

<u>In sections 2.2 and 2.3, only rules in L are considered.</u>

<u>Theorem 1:</u> If $\delta \epsilon L$, then the Bayes risk of δ for given μ' is

$$B(\mu', \delta) = \text{tr } B^T K B \Sigma_1 + \text{tr } \bar{B}^T K \bar{B} \Sigma_0 + (\Delta - \mu')^T \bar{B}^T K \bar{B} (\Delta - \mu') . \tag{7}$$

Proof:

$$B(\mu',\delta) = E(\delta(X) - \bar{\theta})^T K(\delta(X) - \bar{\theta})$$
$$= E[B(X - \bar{\theta}) + \bar{B}(\Delta - \bar{\theta})]^T K[B(X - \bar{\theta}) + \bar{B}(\Delta - \bar{\theta})]$$
$$= \operatorname{tr} B^T K B \Sigma_1 + \operatorname{tr} \bar{B}^T K \bar{B} \Sigma_0 + (\Delta - \mu')^T \bar{B}^T K \bar{B}(\Delta - \mu') \; .$$

Write

$$\mu = \Delta - \mu'$$

and note that, since Δ is fixed, δ is determined by B, and μ' is determined by μ. Thus identify

$$R(\mu,B) = B(\mu',\delta) \; ,$$

and (7) becomes

$$R(\mu,B) = \operatorname{tr} B^T K B \Sigma_1 + \operatorname{tr} \bar{B}^T K \bar{B} \Sigma_0 + \mu^T \bar{B}^T K \bar{B} \mu \; . \tag{8}$$

Similarly, the set U' of (4) becomes

$$U = \{\mu \in E^p \,\big|\, |\mu| \leq M\} \; . \tag{9}$$

By definition, if $\delta_o(x) = B_o x + \bar{B}_o \Delta$ is a Λ-minimax rule in L, then with $\mathcal{T}_p$ the space of real p X p matrices ($= E^{p^2}$),

$$\sup_{\mu \in U} R(\mu,B_o) = \inf_{B \in \mathcal{T}_p} \sup_{\mu \in U} R(\mu,B) \; ,$$

in which case, if there is no risk of confusion, B_o will be called U-minimax in L.

To demonstrate the existence of a U-minimax rule in L and for other reasons (made explicit following theorem 4), we wish to show that a minimax theorem holds. To do so, it will first be shown that in searching for U-minimax rules, attention may be restricted to sets $\beta \subset \mathcal{T}_p$ and $\mathcal{E} \subset U$ for which the following minimax theorem applies. See Stein [14, pp. I.3.1 - I.3.2].

Theorem: (A Minimax Theorem) Let $\mathcal{E}$ be a finite set, β an arbitrary convex set, and R a bounded real valued function on $\mathcal{E} \times \beta$ which is convex in its second argument. Let $\tilde{\mathcal{E}}$ be the set of all probability functions Π on $\mathcal{E}$ and extend the definition of R to $\tilde{\mathcal{E}} \times \beta$ by

$$R(\Pi,B) = \sum_{e \in \mathcal{E}} R(e,B)\Pi(e) . \tag{10}$$

Then

$$\sup_{\Pi \in \tilde{\mathcal{E}}} \inf_{B \in \beta} R(\Pi,B) = \inf_{B \in \beta} \sup_{\Pi \in \tilde{\mathcal{E}}} R(\Pi,B) .$$

From (8) note that $R(\mu,I) = \operatorname{tr} K\Sigma_1 = \tau$ say. Define

$$\beta = \{B \in \mathcal{T}_p \mid \sup_{\mu \in U} R(\mu,B) \leq \tau\} . \tag{11}$$

Now $I \in \beta$ so β is not empty.

Theorem 2: If $B^* \notin \beta$, then B^* is not U-minimax in L.

Proof: Since $B^* \notin \beta$, there exists a $\mu^* \in U$ such that $R(\mu^*,B^*) > \tau$. Thus

$$\inf_{B \in \beta} \sup_{\mu \in U} R(\mu,B) \leq \sup_{\mu \in U} R(\mu,I) = \tau < R(\mu^*,B^*) \leq \sup_{\mu \in U} R(\mu,B^*) .$$

Corollary 2.1:

$$\inf_{B \in \beta} \sup_{\mu \in U} R(\mu,B) = \inf_{B \in \mathcal{T}_p} \sup_{\mu \in U} R(\mu,B) .$$

Thus only $B \in \beta$ need be considered.

We shall need the following result. See, for example, Hadley [4, p. 91].

Theorem: (An Extreme Point Theorem) If f is a convex and continuous function on a (non-empty) convex, compact set $S \subset E^p$, then f assumes its maximum on S at an extreme point of S. If f is strictly convex, the only maxima occur at extreme points.

Let $\mathcal{E} \subset U$ be the collection of extreme points e_i of the convex, compact set U. That is

$$\mathcal{E} = \{e_1, e_2, \cdots, e_u\} = \{\mu \in E^p \,\big|\, |\mu| = M\} . \tag{12}$$

(Note that the number of elements in $\mathcal{E}$ is $u \leq 2^p$, with equality if all $m_j > 0$.) Now for any B, $\bar{B}^T K \bar{B}$ is non-negative definite and thus $\mu^T \bar{B}^T K \bar{B} \mu$ is convex on U. Thus by the extreme point theorem,

$$\sup_{\mu \in U} \mu^T \bar{B}^T K \bar{B} \mu = \sup_{e \in \mathcal{E}} e^T \bar{B}^T K \bar{B} e$$

and so from (8)

$$\sup_{\mu \in U} R(\mu, B) = \sup_{e \in \mathcal{E}} R(e, B) . \tag{13}$$

Lemma 3: The set β is convex, and for each $e \in \mathcal{E}$, $R(e,B)$ is convex in B. A proof is given in the appendix.

Also, for all $e \in \mathcal{E}$, $B \in \beta$,

$$R(e,B) \leq \tau < \infty \tag{14}$$

and so $R(\cdot,\cdot)$ is bounded on $\mathcal{E} \times \beta$.

Define $\tilde{\mathcal{E}}$ to be the convex set of all probability distributions on $\mathcal{E}$:

$$\tilde{\mathcal{E}} = \left\{\Pi = (\pi_1, \pi_2, \cdots, \pi_u) \in E^u \,|\, \pi_i \geq 0,\ i = 1, 2, \cdots, u;\ \sum_{i=1}^{u} \pi_i = 1\right\}, \tag{15}$$

and extend the definition of R to $\tilde{\mathcal{E}} \times \beta$ by

$$R(\Pi,B) = \sum_{i=1}^{u} R(e_i,B)\pi_i \ .$$

By definition, $\mathcal{E}$ is finite, by (14) R is bounded on $\mathcal{E} \times \beta$, and by lemma 3, β is convex and R is convex in its second argument. The hypotheses of the minimax theorem are thus satisfied and so

Theorem 4:

$$\sup_{\Pi\in\tilde{\mathcal{E}}} \inf_{B\in\beta} R(\Pi,B) = \inf_{B\in\beta} \sup_{\Pi\in\tilde{\mathcal{E}}} R(\Pi,B) \ .$$

Theorem 4 guarantees the existence of a saddle point of R; that is, a pair $(\Pi_o,B_o) \in \tilde{\mathcal{E}} \times \beta$ such that for all $(\Pi,B) \in \tilde{\mathcal{E}} \times \beta$,

$$R(\Pi,B_o) \leq R(\Pi_o,B_o) \leq R(\Pi_o,B) \ .$$

See Karlin [7, Vol. II, p. 9]. The theorem thus guarantees the existence of a rule B_o, which is $\tilde{\mathcal{E}}$-minimax in β; that is, such that

$$\sup_{\Pi\in\tilde{\mathcal{E}}} R(\Pi,B_o) = \inf_{B\in\beta} \sup_{\Pi\in\tilde{\mathcal{E}}} R(\Pi,B) \ .$$

Lemma 5 shows that for any $B \in \mathcal{T}_p$,

$$\sup_{\Pi\in\tilde{\mathcal{E}}} R(\Pi,B) = \sup_{\mu\in U} R(\mu,B) \ ,$$

so that

$$\sup_{\mu\in U} R(\mu,B_o) = \inf_{B\in\beta} \sup_{\mu\in U} R(\mu,B)$$

and B_o is U-minimax in β. Finally, by corollary 2.1,

$$\sup_{\mu \in U} R(\mu, B_o) = \inf_{B \in \mathcal{T}_p} \sup_{\mu \in U} R(\mu, B) ;$$

that is, B_o is U-minimax in L.

Note too that theorem 4 simplifies the calculation of a U-minimax rule, for to compute a U-minimax rule directly is to seek B_o such that

$$\sup_{\mu \in U} R(\mu, B_o) = \inf_{B \in \mathcal{T}_p} \sup_{\mu \in U} R(\mu, B) .$$

For each $B \in \mathcal{T}_p$, $R(\cdot, B)$ is convex and continuous on U, and U is compact and convex. Thus by the extreme point theorem, for each $B \in \mathcal{T}_p$, the supremum, $\sup_{\mu \in U} R(\mu, B)$ is attained at an extreme point, $e(B)$ of U; that is, at an element of $\mathcal{E} = \{e_1, e_2, \cdots, e_u\}$.

Let

$$\beta_j = \{B \in \mathcal{T}_p | R(e_j, B) = \sup_{e \in \mathcal{E}} R(e, B)\} ,$$

so that $\mathcal{T}_p = \bigcup_{j=1}^{u} \beta_j$. Now determine $\inf_{B \in \beta_j} R(e_j, B)$ and then $\inf_{1 \le j \le u} \inf_{B \in \beta_j} R(e_j, B)$ to obtain the U-minimax risk. But the minimization over β_j is a constrained minimization; namely

$$\underset{B \in \mathcal{T}_p}{\text{minimize}}\; R(e_j, B)$$

subject to

$$R(e_j, B) = \sup_{e \in \mathcal{E}} R(e, B) = R(e(B), B) .$$

The constraint is complicated in that the relation between $e(B)$ and B is not a simple one.

As will be seen (lemma 6), theorem 4 allows the minimization (for each $\Pi \in \tilde{\mathcal{E}}$) of $R(\Pi, B)$ over all of $\mathcal{T}_p$. This minimization is unconstrained and affords an analytical solution for the minimum by standard techniques of the calculus. Then it remains only to maximize $\inf_{B\in\mathcal{T}_p} R(\Pi,B)$ over the convex, compact set $\tilde{\mathcal{E}}$; a task which is accomplished numerically with relative ease. Some examples of this computation are given in section 2.5 for the case p = 2.

We now return to the derivation of the U-minimax rule.

Lemma 5: For each $B \in \mathcal{T}_p$,

$$\sup_{\mu\in U} R(\mu,B) = \sup_{e\in\mathcal{E}} R(e,B) = \sup_{\Pi\in\tilde{\mathcal{E}}} R(\Pi,B) \ .$$

Proof: Recall that $\mathcal{E} = \{e_1, e_2, \cdots, e_u\}$. Since the distribution Π^i, which assigns probability one to e_i, is in $\tilde{\mathcal{E}}$ for i = 1, 2, $\cdots$, u,

$$\sup_{e\in\mathcal{E}} R(e,B) \le \sup_{\Pi\in\tilde{\mathcal{E}}} R(\Pi,B) \ .$$

Also, for any $\Pi \in \tilde{\mathcal{E}}$,

$$R(\Pi,B) = \sum_{i=1}^{u} R(e_i,B)\pi_i \le \sum_{i=1}^{u} \sup_{e\in\mathcal{E}} R(e,B)\pi_i = \sup_{e\in\mathcal{E}} R(e,B) \ ,$$

and thus

$$\sup_{\Pi\in\tilde{\mathcal{E}}} R(\Pi,B) \le \sup_{e\in\mathcal{E}} R(e,B) \ .$$

The first equality is (13).

2.3 Reduction to a Mathematical Programming Problem

In this section the search for a U-minimax rule will be reduced to a problem of maximizing a continuous function on a compact, convex set---a problem which can be solved by numerical techniques.

Recall that the objective is to find a matrix B_o such that

$$\sup_{\mu\epsilon U} R(\mu,B_o) = \inf_{B\epsilon \mathcal{T}_p} \sup_{\mu\epsilon U} R(\mu,B) .$$

Lemma 6: The U-minimax risk (in L) satisfies

$$\inf_{B\epsilon \mathcal{T}_p} \sup_{\mu\epsilon U} R(\mu,B) = \sup_{\Pi\epsilon\tilde{\mathcal{E}}} \inf_{B\epsilon \mathcal{T}_p} R(\Pi,B) .$$

A proof is given in the appendix.

The next theorem accomplishes the reduction to a maximization problem and, together with theorem 9 of section 2.4, constitutes the main result of this paper.

Theorem 7: In L, the U-minimax risk

$$\inf_{B\epsilon \mathcal{T}_p} \sup_{\mu\epsilon U} R(\mu,B) = \sup_{\Pi\epsilon\tilde{\mathcal{E}}} R(\Pi,B_\Pi)$$

$$= \sup_{\Pi\epsilon\tilde{\mathcal{E}}} \operatorname{tr} K\bar{B}_\Pi(\Sigma_0 + D_\Pi)$$

where

$$B_\Pi = (\Sigma_0 + D_\Pi)(\Sigma_1 + \Sigma_0 + D_\Pi)^{-1} ,$$

and

$$D_\Pi = \sum_{i=1}^{u} \pi_i e_i e_i^T .$$

Furthermore, there is a $\Pi_o \epsilon \tilde{\mathcal{E}}$ such that

$$R(\Pi_o, B_{\Pi_o}) = \sup_{\Pi\epsilon\tilde{\mathcal{E}}} R(\Pi,B_\Pi) ,$$

and the rule δ_o defined by

$$\delta_o(x) = B_{\Pi_o} x + \bar{B}_{\Pi_o} \Delta$$

is U-minimax in L.

Proof: The existence of a U-minimax rule is demonstrated in the comments following theorem 4. Now by lemma 6,

$$\inf_{B\in\mathcal{T}_p} \sup_{\mu\in U} R(\mu,B) = \sup_{\Pi\in\tilde{\mathcal{E}}} \inf_{B\in\mathcal{T}_p} R(\Pi,B) ,$$

and from (8)

$$R(\mu,B) = \text{tr } B^T K B\Sigma_1 + \text{tr } \bar{B}^T K\bar{B}\Sigma_0 + \mu^T\bar{B}^T K\bar{B}\mu$$

so that

$$R(\Pi,B) = \sum_{i=1}^{u} R(e_i,B)\pi_i$$

$$= \text{tr } B^T K B\Sigma_1 + \text{tr } \bar{B}^T K\bar{B}\Sigma_0 + \sum_{i=1}^{u} e_i^T\bar{B}^T K\bar{B}e_i\pi_i \ .$$

Now $e_i^T\bar{B}^T K\bar{B}e_i$ is a scalar, and so

$$e_i^T\bar{B}^T K\bar{B}e_i = \text{tr } e_i^T\bar{B}^T K\bar{B}e_i = \text{tr } \bar{B}^T K\bar{B}e_i e_i^T \ .$$

Thus

$$\begin{aligned} R(\Pi,B) &= \text{tr } [B^T K B\Sigma_1 + \bar{B}^T K\bar{B}(\Sigma_0 + D_\Pi)] \\ &= \text{tr } K[B\Sigma_1 B^T + \bar{B}\Sigma\bar{B}^T] \end{aligned} \qquad (16)$$

by the cyclical property of the trace, where

$$\Sigma = \Sigma_0 + D_\Pi \ .$$

To obtain $B \in \mathcal{T}_p$ to minimize $R(\Pi,B)$, let

$$B = \Sigma(\Sigma_1 + \Sigma)^{-1} + G \ .$$

Then from (16) it can be shown with some computation that

$$R(\Pi,B) = \mathrm{tr}\ K[\Sigma - \Sigma(\Sigma_1 + \Sigma)^{-1}\Sigma] + \mathrm{tr}\ KG(\Sigma_1 + \Sigma)G^T .$$

Since K and $G(\Sigma_1 + \Sigma)G^T$ are at least non-negative definite, all their characteristic roots are non-negative, and therefore all the characteristic roots of $KG(\Sigma_1 + \Sigma)G^T$ are non-negative (Anderson and Gupta [1, Corollary 2.2.1]). Thus

$$\mathrm{tr}\ KG(\Sigma_1 + \Sigma)G^T \geq 0$$

with equality if $G = 0$. It follows that for each $\Pi \in \tilde{\mathcal{E}}$

$$B_\Pi = \Sigma(\Sigma_1 + \Sigma)^{-1} = (\Sigma_0 + D_\Pi)(\Sigma_1 + \Sigma_0 + D_\Pi)^{-1} \tag{17}$$

minimizes $R(\Pi,B)$ over $\mathcal{T}_p$. Finally, with some manipulation

$$\begin{aligned}\inf_{B\in\mathcal{T}_p} R(\Pi,B) &= R(\Pi,B_\Pi)\\ &= \mathrm{tr}\ B_\Pi^T K B_\Pi \Sigma_1 + \mathrm{tr}\ \bar{B}_\Pi^T K \bar{B}_\Pi(\Sigma_0 + D_\Pi)\\ &= \mathrm{tr}\ K\bar{B}_\Pi(\Sigma_0 + D_\Pi) .\end{aligned}$$

2.4 Linear Estimators

In section 2.3, a U-minimax rule in L (6) was derived. It was asserted in section 2.2 that only such rules need be considered; that is, in the class $D_L = \{\ \delta \in D\ |\delta(x) = Bx + C\}$ of all linear rules, any U-minimax rule must be in L. Theorem 9 justifies that assertion.

Lemma 8: For p-vectors $G \neq 0$, μ, and $M \geq 0$, $p \times p$ matrix $\bar{B}$ and positive definite $p \times p$ matrix K,

$$\underset{|\mu| \leq M}{\mathrm{supremum}}[\bar{B}\mu + G]^T K[\bar{B}\mu + G] > \underset{|\mu| \leq M}{\mathrm{supremum}}[\bar{B}\mu]^T K[\bar{B}\mu] . \tag{18}$$

A proof is given in the appendix.

Theorem 9: If the rule $\delta(x) = Bx + C$ is U-minimax among all linear rules, then $C = \bar{B}\Delta$, so that U-minimax rules in D_L are of the form $\delta_o(x) = Bx + \bar{B}\Delta$; that is, $\delta \in L$.

Proof: Without loss of generality, write $\delta(x) = Bx + \bar{B}\Delta + G$, where G is the p-vector $C - \bar{B}\Delta$. From (5), the Bayes risk of δ for given μ' is

$$B(\mu',\delta) = E(\delta(X) - \underline{\bar{\theta}})^T K(\delta(X) - \underline{\bar{\theta}})$$

$$= \operatorname{tr} B^T K B \Sigma_1 + \operatorname{tr} \bar{B}^T K \bar{B} \Sigma_0 + [\bar{B}(\Delta-\mu') + G]^T K[\bar{B}(\Delta-\mu') + G] .$$

Since $\delta_o(x)$ is $\delta(x)$ with $G = 0$,

$$B(\mu',\delta_o) = \operatorname{tr} B^T K B \Sigma_1 + \operatorname{tr} \bar{B}^T K \bar{B} \Sigma_0 + [\bar{B}(\Delta-\mu')]^T K[\bar{B}(\Delta-\mu')] .$$

It must be shown that if $G \neq 0$,

$$\sup_{\mu' \in U'} B(\mu',\delta) > \sup_{\mu' \in U'} B(\mu',\delta_o) . \tag{19}$$

But this is an immediate consequence of lemma 8.

Notice that in sections 2.2 and 2.3 the non-negative definite character of K was used (lemma 3, theorem 7), but that only in this section (lemma 8) was it necessary to require K to be positive definite. If it is assumed that K is non-negative definite and δ_1 is U-minimax in D_L, there exists a rule δ_o in L with U-minimax risk not exceeding that of δ_1. This is seen by noting that lemma 8 holds for K non-negative definite if ">" is replaced by "≥", and thus (19) holds subject to the same substitution.

As an application, let $W = (w_1, w_2, \cdots, w_p)^T$ be a vector of non-zero constants, and suppose

$$W^T \underline{\bar{\theta}} = \sum_{i=1}^{p} w_i \underline{\bar{\theta}}_i$$

is to be estimated. It is proposed to estimate $W^T\bar{\theta}$ by $W^{*T}\delta$ where $\delta \in D_L$, and $W^* = (w_1^*, w_2^*, \cdots, w_p^*)^T$.

Taking as loss function

$$L(W^*,\delta,\theta,x,\mu) = \frac{(W^{*T}\delta(x) - W^T\theta)^2}{W^TW} ,$$

it can be shown that the appropriate estimator of $W^T\bar{\theta}$ is $W^T\gamma(X)$ where γ is a Λ-minimax estimator of $\bar{\theta}$ corresponding to the non-negative definite loss matrix $K = \dfrac{WW^T}{W^TW}$.

Note that the previous results apply if the set U of (9) is an arbitrary closed convex polyhedron with extreme points $\{e_i\}$. This is true because the crucial assumption in the minimax theorem of section 2.2 is the finiteness of the number of extreme points.

Observe too that, appropriately modified, the techniques are applicable to compact, convex sets other than convex polyhedra. For example, suppose

$$U = \{\mu \in E^p | \mu^T G\mu \le m\}$$

is an ellipsoid in p-dimensional Euclidean space. Then, by the extreme point theorem, it still follows that the maximum over U of the Bayes risk occurs at an extreme point of U. A complication arises in that the set of extreme points,

$$\mathcal{E} = \{\mu \in E^p | \mu^T G\mu = m\} ,$$

is the boundary of the ellipsoid and is not finite. Thus the minimax theorem does not apply. However, a more general version of the theorem only requires the set of extreme points to be compact in the Wald topology. For a precise statement see Stein [14, p. I.3.7].

2.5 Example: Estimation of a Bivariate Location Parameter

This section presents numerical examples for the case p = 2. With $m_1, m_2 > 0$, the notation becomes:

$$K = \begin{bmatrix} k_{11} & k_{12} \\ k_{12} & k_{22} \end{bmatrix}, \quad \Sigma_0 = \begin{bmatrix} \sigma_{11}^{o} & \sigma_{12}^{o} \\ \sigma_{12}^{o} & \sigma_{22}^{o} \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} \sigma_{11}^{1} & \sigma_{12}^{1} \\ \sigma_{12}^{1} & \sigma_{22}^{1} \end{bmatrix},$$

$$M = (m_1, m_2)^T, \quad \Pi = (\pi_1, \pi_2, \pi_3, \pi_4)^T .$$

Also,

$$e_1 = M, \quad e_2 = (-m_1, m_2)^T, \quad e_3 = -e_2, \quad e_4 = -e_1,$$

and

$$D_q = \sum_{i=1}^{4} \pi_i e_i e_i^T = \begin{bmatrix} m_1^2 & q m_1 m_2 \\ q m_1 m_2 & m_2^2 \end{bmatrix},$$

where $q = (\pi_1 + \pi_4) - (\pi_2 + \pi_3)$ and $q \in [-1,1]$. Note that the maximizing Π depends only on the difference, q, in weights placed on the diagonals of the box. The computations are based on theorem 7 which becomes:

Corollary 7.1: If p = 2, the U-minimax risk of the linear U-minimax rule is given by

$$R(B_o) \equiv \underset{q \in [-1,1]}{\text{supremum}} \ \text{tr}\ K\Sigma_1(\Sigma_1 + \Sigma_0 + D_q)^{-1}(\Sigma_0 + D_q) ,$$

and if the supremum is attained at $q = q_o$, then the rule defined by

$$\delta_o(x) = B_o x + \bar{B}_o \Delta ,$$

with

$$B_o = (\Sigma_0 + D_{q_o})(\Sigma_1 + \Sigma_0 + D_{q_o})^{-1}$$

is U-minimax.

For various choices of K, Σ_0, Σ_1, and M, Table 1 gives the U-minimax rule, its U-minimax risk, and the maximizing q_o, together with the corresponding (linear) Bayes rule and its risk. A (linear) Bayes rule (with respect to the prior distribution λ) is a rule $\delta_1 \epsilon D_L$ such that

$$E_\lambda\{E[(\delta_1(X) - \underline{\bar{\theta}})^T K(\delta_1(X) - \underline{\bar{\theta}})|\underline{\bar{\theta}}]\} = \inf_{\delta \epsilon D_L} E_\lambda\{E[(\delta(X) - \underline{\bar{\theta}})^T K(\delta(X) - \underline{\bar{\theta}})|\underline{\bar{\theta}}]\} .$$

In what follows, B_1 is the coefficient matrix of x in the (linear) Bayes rule

$$\delta_1(x) = B_1 x + \bar{B}_1 \mu .$$

That is, $B_1 = \Sigma_0(\Sigma_1 + \Sigma_0)^{-1}$. The Bayes risk of the (linear) Bayes rule is

$$R(B_1) \equiv \text{tr } K\Sigma_1 B_1^T .$$

The first three examples illustrate the effect of changing K. There is no effect on the Bayes rule (only on its Bayes risk), but the U-minimax rule does change. Examples 4 and 5; 6, 7, and 8; and 14, 15, and 16, show the effect of changing M. Examples 6 and 9; and 10, 11, and 12, have different Σ_1, while 6, 12, 13, and 14, differ in Σ_0. Examples 6, 12, and 14 show that B_o can be diagonal when B_1 is not. Notice that the examples show many values of q_o.

Table 1.--Selected results for the location problem with p = 2

Example	K	Σ_0	Σ_1	M	q_0	B_0	$R(B_0)$	B_1	$R(B_1)$
1	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}1\\ 1\end{bmatrix}$	0.00	$\begin{bmatrix}.667 & 0\\ 0 & .667\end{bmatrix}$	1.333	$\begin{bmatrix}.500 & 0\\ 0 & .500\end{bmatrix}$	1.000
2	$\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}$	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}1\\ 1\end{bmatrix}$	0.80	$\begin{bmatrix}.641 & .096\\ .096 & .641\end{bmatrix}$	2.756	$\begin{bmatrix}.500 & 0\\ 0 & .500\end{bmatrix}$	2.000
3	$\begin{bmatrix}2 & 1\\ 1 & 3\end{bmatrix}$	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}1\\ 1\end{bmatrix}$	0.63	$\begin{bmatrix}.651 & .073\\ .073 & .651\end{bmatrix}$	3.403	$\begin{bmatrix}.500 & 0\\ 0 & .500\end{bmatrix}$	2.500
4	$\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}$	$\begin{bmatrix}2 & 1\\ 1 & 5\end{bmatrix}$	$\begin{bmatrix}3 & 0\\ 0 & 2\end{bmatrix}$	$\begin{bmatrix}1\\ 1\end{bmatrix}$	0.58	$\begin{bmatrix}.473 & .104\\ .069 & .736\end{bmatrix}$	6.197	$\begin{bmatrix}.382 & .088\\ .059 & .706\end{bmatrix}$	5.471
5	$\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}$	$\begin{bmatrix}2 & 1\\ 1 & 5\end{bmatrix}$	$\begin{bmatrix}3 & 0\\ 0 & 2\end{bmatrix}$	$\begin{bmatrix}2\\ 1\end{bmatrix}$	0.56	$\begin{bmatrix}.644 & .094\\ .063 & .733\end{bmatrix}$	7.177	$\begin{bmatrix}.382 & .088\\ .059 & .706\end{bmatrix}$	5.471
6	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\\ 1 & 5\end{bmatrix}$	$\begin{bmatrix}3 & 0\\ 0 & 2\end{bmatrix}$	$\begin{bmatrix}1\\ 1\end{bmatrix}$	-1.00	$\begin{bmatrix}.500 & 0\\ 0 & .750\end{bmatrix}$	3.000	$\begin{bmatrix}.382 & .088\\ .059 & .706\end{bmatrix}$	2.559
7	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\\ 1 & 5\end{bmatrix}$	$\begin{bmatrix}3 & 0\\ 0 & 2\end{bmatrix}$	$\begin{bmatrix}2\\ 1\end{bmatrix}$	-0.50	$\begin{bmatrix}.667 & 0\\ 0 & .750\end{bmatrix}$	3.500	$\begin{bmatrix}.382 & .088\\ .059 & .706\end{bmatrix}$	2.559
8	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\\ 1 & 5\end{bmatrix}$	$\begin{bmatrix}3 & 0\\ 0 & 2\end{bmatrix}$	$\begin{bmatrix}.1\\ .1\end{bmatrix}$	-1.00	$\begin{bmatrix}.384 & .087\\ .058 & .707\end{bmatrix}$	2.565	$\begin{bmatrix}.382 & .088\\ .059 & .706\end{bmatrix}$	2.559
9	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\\ 1 & 5\end{bmatrix}$	$\begin{bmatrix}3 & 2\\ 2 & 2\end{bmatrix}$	$\begin{bmatrix}1\\ 1\end{bmatrix}$	1.00	$\begin{bmatrix}.500 & 0\\ -.250 & .875\end{bmatrix}$	2.750	$\begin{bmatrix}.423 & -.038\\ -.308 & .846\end{bmatrix}$	2.269
10	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}$	$\begin{bmatrix}1\\ 1\end{bmatrix}$	0.63	$\begin{bmatrix}.583 & .032\\ -.046 & .707\end{bmatrix}$	1.859	$\begin{bmatrix}.500 & 0\\ 0 & .500\end{bmatrix}$	1.500

Table 1.--Continued

Example	K	Σ_0	Σ_1	M	q_0	B_0	$R(B_0)$	B_1	$R(B_1)$
11	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}$	$\begin{bmatrix}3 & -1\\-1 & 2\end{bmatrix}$	$\begin{bmatrix}1\\1\end{bmatrix}$	-1.00	$\begin{bmatrix}.522 & .130\\.087 & .522\end{bmatrix}$	2.391	$\begin{bmatrix}.400 & .333\\.200 & .333\end{bmatrix}$	1.333
12	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}$	$\begin{bmatrix}3 & 0\\0 & 2\end{bmatrix}$	$\begin{bmatrix}1\\1\end{bmatrix}$	-1.00	$\begin{bmatrix}.500 & 0\\0 & .500\end{bmatrix}$	2.500	$\begin{bmatrix}.357 & .214\\.143 & .286\end{bmatrix}$	1.640
13	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 0\\0 & 4\end{bmatrix}$	$\begin{bmatrix}3 & 0\\0 & 2\end{bmatrix}$	$\begin{bmatrix}1\\1\end{bmatrix}$	0.00	$\begin{bmatrix}.500 & 0\\0 & .714\end{bmatrix}$	2.929	$\begin{bmatrix}.400 & 0\\0 & .667\end{bmatrix}$	2.533
14	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	$\begin{bmatrix}5 & 1\\1 & 2\end{bmatrix}$	$\begin{bmatrix}3 & 0\\0 & 2\end{bmatrix}$	$\begin{bmatrix}1\\1\end{bmatrix}$	-1.00	$\begin{bmatrix}.667 & 0\\0 & .600\end{bmatrix}$	3.200	$\begin{bmatrix}.613 & .097\\.065 & .484\end{bmatrix}$	2.806
15	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	$\begin{bmatrix}5 & 1\\1 & 2\end{bmatrix}$	$\begin{bmatrix}3 & 0\\0 & 2\end{bmatrix}$	$\begin{bmatrix}2\\1\end{bmatrix}$	-0.50	$\begin{bmatrix}.750 & 0\\0 & .600\end{bmatrix}$	3.450	$\begin{bmatrix}.613 & .097\\.065 & .484\end{bmatrix}$	2.806
16	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	$\begin{bmatrix}5 & 1\\1 & 2\end{bmatrix}$	$\begin{bmatrix}3 & 0\\0 & 2\end{bmatrix}$	$\begin{bmatrix}2\\2\end{bmatrix}$	-0.25	$\begin{bmatrix}.750 & 0\\0 & .750\end{bmatrix}$	3.750	$\begin{bmatrix}.613 & .097\\.065 & .484\end{bmatrix}$	2.806

3. A SPECIAL CASE OF THE LOCATION PROBLEM

If $M = (0, 0, \cdots, 0)^T$, so that $\mu' = \Delta$ is known, the Λ-minimax rule in D_L reduces to the (linear) Bayes rule. For if $M = 0$, the set $\mathcal{E}$ consists of only the p-vector zero, and $\tilde{\mathcal{E}} = \{\Pi\}$ where Π assigns probability one to the vector zero. Thus D_Π of theorem 7 is zero, and the Λ-minimax rule is

$$\delta_o(x) = B_o x + \bar{B}_o \mu ,$$

where

$$B_o = \Sigma_0(\Sigma_1 + \Sigma_0)^{-1} .$$

This can be shown to be the (linear) Bayes rule. The result does not depend on the loss matrix K. Note that the (linear) Bayes rule has the coefficient matrix B_o diagonal whenever Σ_0 and Σ_1 are diagonal. This seems reasonable since if the X_i and $\bar{\theta}_j$ are uncorrelated, one would not expect to obtain information about $\bar{\theta}_j$ from X_i $(i \neq j)$. It will now be shown that if the loss matrix K is diagonal, the Λ-minimax rule in D_L also has this property. The result in this case is an analytic expression for the Λ-minimax rule in D_L. The derivation will also serve to illustrate a technique which is often useful in minimax theory.

Suppose $\Sigma_1 = \text{diag.}(\sigma_i^1)$, $\Sigma_0 = \text{diag.}(\sigma_i^o)$, and $K = \text{diag.}(k_i)$. Recall that B_o is sought such that

$$\sup_{\mu \in U} R(\mu, B_o) = \inf_{B \in \mathcal{T}_p} \sup_{\mu \in U} R(\mu, B)$$

or equivalently, by (13), such that

$$\sup_{e \in \mathcal{E}} R(e, B_o) = \inf_{B \in \mathcal{T}_p} \sup_{e \in \mathcal{E}} R(e, B) .$$

Define

$$B_o = \text{diag.}\left[\frac{\sigma_i^o + m_i^2}{\sigma_i^1 + \sigma_i^o + m_i^2} \right] . \tag{20}$$

Needing only to consider rules in L, a distribution λ on $\mathcal{E}$ will be found such that B_o minimizes $E_\lambda R(e,B)$ and it will be shown that $R(e,B_o)$ is constant on $\mathcal{E}$. It will follow that δ_o defined by

$$\delta_o(x) = B_o x + \bar{B}_o \Delta$$

with B_o as in (20), is Λ-minimax in D_L (theorem 12).

<u>Lemma 10</u>: With B_o diagonal and with Σ_0, Σ_1, and K diagonal, $R(e,B_o) = R$ is constant on $\mathcal{E}$.

<u>Proof</u>: Notice from (8) that for an arbitrary element $e_i \in \mathcal{E}$,

$$\begin{aligned} R(e_i,B_o) &= \operatorname{tr} B_o^T K B_o \Sigma_1 + \operatorname{tr} \bar{B}_o^T K \bar{B}_o \Sigma_0 + \operatorname{tr} \bar{B}_o^T K \bar{B}_o e_i e_i^T \\ &= \operatorname{tr} B_o^T K B_o \Sigma_1 + \operatorname{tr} \bar{B}_o^T K \bar{B}_o \Sigma_0 + \sum_{j=1}^{p} (\bar{B}_o^T K \bar{B}_o)_{jj} m_j^2 \\ &= R \text{ say,} \end{aligned}$$

since $\bar{B}_o^T K \bar{B}_o$ is diagonal, and the diagonal elements of $e_i e_i^T$ are m_1^2, m_2^2, $\cdots$, m_p^2 for every $i = 1, 2, \cdots, u$.

<u>Lemma 11</u>: With B_o as in (20), there exists a distribution λ_o on $\mathcal{E}$ such that

$$R = E_{\lambda_o} R(e,B_o) = \inf_{B \in \mathcal{T}_p} E_{\lambda_o} R(e,B) .$$

A proof is given in the appendix.

<u>Theorem 12</u>: With $\Sigma_0 = \text{diag.}(\sigma_i^0)$, $\Sigma_1 = \text{diag.}(\sigma_i^1)$, and K diagonal, the Λ-minimax rule in D_L is given by

$$\delta_o(x) = B_o x + \bar{B}_o \Delta ,$$

where B_o is defined in (20).

Proof: With λ_o as in lemma 11, for any $B \in \mathcal{T}_p$ observe that

$$\sup_{e\in\mathcal{E}} R(e,B_o) = \sum_{i=1}^{u} R\lambda_o(e_i) \qquad \text{by lemma 10}$$

$$= \sum_{i=1}^{u} R(e_i,B_o)\lambda_o(e_i)$$

$$\leq \sum_{i=1}^{u} R(e_i,B)\lambda_o(e_i) \qquad \text{by lemma 11}$$

$$\leq \sup_{e\in\mathcal{E}} R(e,B) .$$

Thus

$$\sup_{e\in\mathcal{E}} R(e,B_o) = \inf_{B\in\mathcal{T}_p} \sup_{e\in\mathcal{E}} R(e,B)$$

so that δ_o is Λ-minimax in D_L.

Corollary 12.1: If $p = 1$, $\Sigma_1 = \sigma_1^2$, $\Sigma_0 = \sigma_o^2$, $K = k$, then the Λ-minimax rule in D_L is

$$\delta_o(x) = \frac{\dfrac{x}{\sigma_1^2} + \dfrac{\Delta}{\sigma_o^2 + M^2}}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_o^2 + M^2}} .$$

with Λ-minimax risk

$$k \left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_o^2 + M^2} \right]^{-1} .$$

4. SOME RELATED RESULTS FOR THE LOCATION PROBLEM

4.1 Other Distribution Assumptions

In considering quadratic form loss functions and linear rules, the first two moments sufficiently characterize the prior and sampling distributions. Subject to this restriction, the assumptions of section 2.1 may be weakened by assuming not that $\mathrm{Cov}(X|\bar{\underset{\sim}{\theta}}) = \Sigma_1$, but only that $\mathrm{Cov}(X|\bar{\underset{\sim}{\theta}}) = \Sigma_1(\bar{\underset{\sim}{\theta}})$ where

$$E\{\Sigma_1(\bar{\underset{\sim}{\theta}})|\mu'\} = \Sigma_1$$

independent of μ'. Phrasing this somewhat differently, suppose that both the sampling distribution's mean $\bar{\underset{\sim}{\theta}}$ and covariance Σ are unknown, and that the decision maker has a joint prior distribution on $\bar{\underset{\sim}{\theta}}$ and Σ such that under this prior, $E(\bar{\underset{\sim}{\theta}}) = \mu'$, $\mathrm{Cov}(\bar{\underset{\sim}{\theta}}) = \Sigma_0$, and $E(\Sigma) = \Sigma_1$, where Σ_0 and Σ_1 are known, and it has been learned that $\mu' \epsilon U'$. The results are identical to those for the original model.

As a further example of changing the distribution assumptions, suppose X is a (univariate) positive random variable with mean $\bar{\underset{\sim}{\theta}}$ and variance $\mathrm{var}(X|\bar{\underset{\sim}{\theta}}) = \nu\bar{\underset{\sim}{\theta}}$, where $\nu > 0$ is known. Again suppose that the prior distribution is such that $E(\bar{\underset{\sim}{\theta}}) = \mu'$, $\mathrm{var}(\bar{\underset{\sim}{\theta}}) = \sigma_0^2$, where σ_0^2 is known and

$$\mu'\epsilon U' = \{\mu'\epsilon E^1 \,\big|\, |\Delta - \mu'| \le M\} \; .$$

Considering rules of the form $\delta(x) = bx + \bar{b}\Delta$ and quadratic loss,

$$B(\mu',\delta) = E(bX + \bar{b}\Delta - \bar{\underset{\sim}{\theta}})^2 = b^2\nu\mu + \bar{b}^2[(\Delta - \mu')^2 + \sigma_0^2] \; .$$

Now

$$\frac{\partial^2 B(\mu',\delta)}{\partial\mu'^2} = 2\bar{b}^2 \ge 0$$

for all μ', so $B(\mu',\delta)$ is convex. It is clearly continuous, so by the extreme point theorem it assumes its maximum at an extreme point $\mu' = \Delta \pm M$. In particular

$$\sup_{\mu'\epsilon U'} B(\mu',\delta) = b^2\nu(\Delta + M) + \bar{b}^2(M^2 + \sigma_0^2) \; .$$

The minimizing b is

$$b = \frac{M^2 + \sigma_0^2}{\nu(\Delta + M) + M^2 + \sigma_0^2}$$

and the U"-minimax risk is

$$R = \left[\frac{1}{\nu(\Delta + M)} + \frac{1}{M^2 + \sigma_0^2} \right]^{-1} .$$

Thus the U'-minimax rule is

$$\delta(x) = \left[\frac{x}{\nu(\Delta + M)} + \frac{\Delta}{M^2 + \sigma_0^2} \right] R .$$

Note that R involves Δ but is bounded above by $M^2 + \sigma_0^2$. Recall that for the original model, the Λ-minimax risk (theorem 7) does not involve Δ.

4.2 Prior Mean Known, Variance Unknown

Finally consider a univariate extension which involves changing both distribution assumptions and the manner in which the prior distribution is specified. Suppose now that the moments of the prior and sampling distributions are as in the original problem (section 2.1), but that μ' is known and $\Sigma_0 = \sigma_0^2$ is not. Assume, however, that the decision maker can assert that

$$0 \le \tau_1 \le \sigma_0^2 \le \tau_2 .$$

Considering rules of the form

$$\delta(x) = bx + \bar{b}\mu' ,$$

maintaining the quadratic loss structure, and performing the now familiar computation, with obvious change of notation, it follows that

$$B(\sigma_o^2,\delta) = E(bX + \bar{b}\mu' - \underline{\theta})^2 = E[b(X - \underline{\theta}) + \bar{b}(\mu' - \underline{\theta})]^2 = b^2\sigma_1^2 + \bar{b}^2\sigma_o^2 \, .$$

Clearly

$$\underset{\tau_1 \le \sigma_o^2 \le \tau_2}{\text{supremum}} \; B(\sigma_o^2,\delta) = b^2\sigma_1^2 + \bar{b}^2\tau_2 \, .$$

The minimizing rule is therefore

$$\delta(x) = \frac{\dfrac{x}{\sigma_1^2} + \dfrac{\mu'}{\tau_2}}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\tau_2}} \, ,$$

and its risk is

$$\left[\frac{1}{\sigma_1^2} + \frac{1}{\tau_2} \right]^{-1} .$$

Since the procedure is minimax, it is not surprising that τ_2, the largest permissible value of σ_o^2 appears as the prior variance.

4.3 Estimation of a Univariate Location Parameter When the Scale Parameter is Unknown

A deterrent to the applicability of the Λ-minimax estimate of location under the model of section 2.1 is the assumption that the variance of the sampling distribution is known. In this section a univariate Normal problem is discussed in which this variance is unknown, but a prior distribution for it is known. The necessary distribution theory can be found in Raiffa and Schlaifer [9, chapters 7 and 11].

Let $X_1, X_2, \cdots, X_n$ be independent identically Normally distributed random variables with mean θ and variance $\frac{1}{h}$ (precision h). Let

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad V = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 ,$$

with the convention that $V = 0$ if $n \leq 1$. The joint probability density function of $\bar{X}$ and V is then

$$f(\bar{x},v|\theta,h) = (\text{const.}) \left[e^{-\frac{1}{2}hn(\bar{x} - \theta)^2} h^{\frac{1}{2}} \right] \left[e^{-\frac{1}{2}hrv} h^{\frac{1}{2}r} \right]$$

where $r = n - 1$.

It is now supposed that the family of natural conjugate prior distributions is rich enough to include an incompleteness specification which is satisfactory for the decision maker. The natural conjugate prior distribution for $\underline{\bar{\theta}}$ and $\tilde{h}$ is the Normal-gamma, with density

$$f_{N_\gamma}(\theta,h|\mu',v',n^*,n') = f_N(\theta|\mu',hn^*)f_{\gamma_2}(h|v',n')$$

$$= (\text{const.}) \left[e^{-\frac{1}{2}hn^*(\theta-\mu')^2} h^{\frac{1}{2}} \right] \left[e^{-\frac{1}{2}hn'v'} h^{\frac{1}{2}n'-1} \right]$$

for

$$-\infty \leq \theta,\mu' \leq \infty, \quad h \geq 0, \quad \text{and } v', n^* > 0, n' > 2 .$$

The following expectations will be needed:

$$E(\bar{X}|\underline{\bar{\theta}},\tilde{h}) = \underline{\bar{\theta}}, \quad E\underline{\bar{\theta}} = \mu', \quad E\left(\frac{1}{\tilde{h}}\right) = \frac{n'v'}{n'-2} \equiv \psi$$

$$\text{var}(\bar{X}|\underline{\bar{\theta}},\tilde{h}) = \frac{1}{\tilde{h}n}, \quad \text{var}\underline{\bar{\theta}} = \frac{v'n'}{n^*(n'-2)} = n^{*-1}\psi . \qquad (21)$$

Suppose that $\underline{\bar{\theta}}$ is to be estimated with quadratic loss, but as before the mean, μ' of the prior distribution on $\underline{\bar{\theta}}$ is unknown. Again assume that the decision maker can assert that

$$\mu' \epsilon U' = \{\mu' \,\big|\, |\Delta - \mu'| \leq M\} ,$$

and let Λ be the set of all Normal-gamma distributions with $\mu' \epsilon U'$.

Considering only rules linear in $\bar{X}$, write

$$\delta(\bar{x}) = b\bar{x} + c\Delta ,$$

where b, c do not depend on v, as an estimate of $\underline{\bar{\theta}}$ when $\bar{x}$ is the observed value of $\bar{X}$. A Λ-minimax rule requires b, c to minimize

$$\sup_{\mu' \epsilon U'} E[b\bar{X} + c\Delta - \underline{\bar{\theta}}]^2 .$$

Now using the expectations in display (21), it can be shown with some computation that

$$E[b\bar{X} + c\Delta - \underline{\bar{\theta}}]^2 = (n^{-1}b^2 + n^{*-1}\bar{b}^2)\psi + \bar{b}^2[(\Delta - \mu') + \frac{c - \bar{b}}{\bar{b}}\Delta]^2 . \qquad (22)$$

The maximum of this expression for $|\Delta - \mu'| \leq M$ is

$$(n^{-1}b^2 + n^{*-1}\bar{b}^2)\psi + \bar{b}^2\left[M + \left|\frac{(c - \bar{b})\Delta}{\bar{b}}\right|\right]^2 ,$$

which in turn is minimized for $c = \bar{b}$. (Even if $\Delta = 0$, so that any c gives the same value of the expression, the rule

$$\delta(\bar{x}) = b\bar{x} + \bar{b}\Delta = b\bar{x} + c\Delta = b\bar{x}$$

for all c.)

It remains to minimize

$$(n^{-1}b^2 + n^{*-1}\bar{b}^2)\psi + \bar{b}^2M^2$$

over values of b. The result is

$$b = \frac{n^{*-1}\psi + M^2}{(n^{-1} + n^{*-1})\psi + M^2}$$

and the Λ-minimax rule is

$$\delta(\bar{x}) = b\bar{x} + c\Delta$$

$$= \frac{\dfrac{\bar{x}}{n^{-1}\psi} + \dfrac{\Delta}{n^{*-1}\psi + M^2}}{\dfrac{1}{n^{-1}\psi} + \dfrac{1}{n^{*-1}\psi + M^2}} \tag{23}$$

or equivalently

$$= \frac{\dfrac{\bar{x}}{\frac{1}{n}E\frac{1}{\tilde{h}}} + \dfrac{\Delta}{\operatorname{var}\bar{\underset{\sim}{\theta}} + M^2}}{\dfrac{1}{\frac{1}{n}E\frac{1}{\tilde{h}}} + \dfrac{1}{\operatorname{var}\bar{\underset{\sim}{\theta}} + M^2}} .$$

Recall that when h is non-random and known, ($\sigma_1^2 = \frac{1}{nh}$ and $\operatorname{var}\bar{\underset{\sim}{\theta}} = \sigma_0^2$), the Λ-minimax rule is (corollary 12.1)

$$\frac{\dfrac{\bar{x}}{\sigma_1^2} + \dfrac{\Delta}{\sigma_0^2 + M^2}}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_0^2 + M^2}} .$$

Observe that this is the rule of display (23) with $E\frac{1}{\tilde{h}}$ replaced by $\frac{1}{h}$.

Knowledge of the parameter n' is not required if the loss function $h(\delta(\bar{x}) - \theta)^2$ is used. The Λ-minimax rule becomes

$$\frac{\dfrac{\bar{x}}{n^{-1}v'} + \dfrac{\Delta}{n^{*-1}v' + M^2}}{\dfrac{1}{n^{-1}v'} + \dfrac{1}{n^{*-1}v' + M^2}} .$$

In (22) the expected loss is a decreasing function of n^*. Thus if the incompleteness in the specification of a prior distribution is to extend to n^*, it must be of the form $n^* \geq n_o$ say, in which case the maximum risk occurs at n_o. The rule with respect to this incompleteness specification would be the rule in (23) with n^* replaced by n_o. Since the precision of the prior distribution for $\bar{\underline{\theta}}$ is hn^*, and h is the precision of the sampling process, this incompleteness specification treats the prior information as equivalent to at least n_o observations on the process. For further comment on equivalent prior samples, see Winkler [17].

5. Λ-MINIMAX ESTIMATES OF SCALE

This section considers Λ-minimax estimation of the variance of a (univariate) Normal distribution when 1) the mean is known and 2) a complete prior distribution is available for the mean. Case 2) is reduced to case 1). Here Λ is a subset of the class of natural conjugate prior distributions. Again, Raiffa and Schlaifer [9, chapters 7 and 11] supplies the distribution theory.

5.1 Mean Known

For the location problem, the distribution assumptions involved only moments and in particular did not assume Normality. Now, suppose that $X_1, X_2, \cdots, X_n$ are independent random variables with identical Normal distributions with mean 0 and variance $h^{-1} = \sigma^2$. Define

$$W = \frac{1}{n}\sum_{i=1}^{n} X_i^2 ,$$

with the convention that W = 0 for n = 0. Then (nh)W has the X^2 distribution with n degrees of freedom.

Suppose σ^2 is to be estimated with loss function

$$\left[\frac{\delta(w) - \sigma^2}{\sigma^2} \right]^2 = h^2(\delta(w) - h^{-1})^2$$

where w, σ^2, and h are realizations of the random variables W, $\tilde{\sigma}^2$, and $\tilde{h}$. Thus a mis-estimation of small values of σ^2 is more costly than an equal mis-estimation of large values.

The natural conjugate prior distribution for $\tilde{h}$ is the gamma-2 with probability density

$$f_{\gamma_2}(h|v',n') = (\text{const.})e^{-\frac{1}{2}hn'v'}h^{\frac{1}{2}n'-1} .$$

The following moments will be required:

$$E(W|\tilde{h}) = \tilde{h}^{-1}, \qquad \text{var}(W|\tilde{h}) = \frac{2}{n\tilde{h}^2}$$
$$E(\tilde{h}) = \frac{1}{v'} , \qquad \text{var}(\tilde{h}) = \frac{2}{n'v'^2} , \tag{24}$$

and thus

$$E(\tilde{h}^2) = \left(\frac{1}{v'} \right)^2 \left(\frac{n' + 2}{n'} \right) .$$

Suppose that the prior distribution is known except for v', and that the decision maker can specify

$$v' \in U = \{v' | \frac{\Delta}{A} \le v' \le \Delta A\}$$

where $\Delta \ge 0$, $A \ge 1$. Then Λ is the class of gamma-2 distributions,

$$\{f_{\gamma_2}(\cdot|v',n'), \ v' \in U\} .$$

As before, considering only estimators which are linear in the sufficient statistic, without loss of generality write

$$\delta(w) = bw + \bar{b}c\varkappa\Delta$$

where $\varkappa = \dfrac{n'}{n' + 2}$ and put $\bar{\varkappa} = 1 - \varkappa$.

For a Λ-minimax rule, b and c must minimize

$$\sup_{v' \in U} E\tilde{h}^2(bW + \bar{b}c\varkappa\Delta - \tilde{h}^{-1})^2 .$$

Using the moments (24), it can be shown that

$$E\tilde{h}^2(bW + \bar{b}c\varkappa\Delta - \tilde{h}^{-1})^2 = \frac{2b^2}{n} + \bar{\varkappa}\,\bar{b}^2 + \varkappa\bar{b}^2[\frac{c\Delta}{v'} - 1]^2 .$$

This is a continuous convex function of $\frac{1}{v'}$, and the set

$$U = \{v' \mid \frac{\Delta}{A} \le v' \le \Delta A\} = \{v' \mid \frac{1}{\Delta A} \le \frac{1}{v'} \le \frac{A}{\Delta}\}$$

is convex and compact so by the extreme point theorem

$$B(b,c) \equiv \sup_{v' \in U} E\,\tilde{h}^2(bW + \bar{b}c\varkappa\Delta - \tilde{h}^{-1})^2$$

$$= \frac{2b^2}{n} + \bar{\varkappa}\,\bar{b}^2 + \varkappa\bar{b}^2 \max\{(\frac{c}{A} - 1)^2, (cA - 1)^2\} .$$

Constants b and c are sought to minimize $B(b,c)$. Let

$$\mathscr{S} = \{c \mid (cA - 1)^2 \ge (\frac{c}{A} - 1)^2\} .$$

Clearly, to minimize $B(b,c)$ subject to $c \in \mathscr{S}$, choose $c = c_o$ such that $(c_oA - 1)^2 = (\frac{c_o}{A} - 1)^2$. On the complement of $\mathscr{S}$, no minimum is attained (the infimum is at c_o). Thus $B(b,c)$ is minimized over all E^1 at the solution, c_o, of $(c_oA - 1)^2 = (\frac{c_o}{A} - 1)^2$. Thus it remains to solve

$$|cA - 1| = |\frac{c}{A} - 1| \qquad (25)$$

for c. If $\frac{c}{A} > 1$, then since $A^2 \ge 1$, $cA = \frac{c}{A}A^2 > 1$ and (25) becomes

$$cA - 1 = \frac{c}{A} - 1 \qquad (26)$$

which has no solution for $A > 1$. If $A = 1$, v' is known, and the problem reduces to the Bayes problem. If $\frac{c}{A} \le 1$ and $cA < 1$, (25) becomes (26) and the same comment applies. Finally, if $\frac{c}{A} \le 1$ but $cA \ge 1$, then (25) becomes

$$cA - 1 = 1 - \frac{c}{A}$$

which has as solution

$$c_o = 2(A + \frac{1}{A})^{-1} = 2\,\frac{A}{A^2 + 1} \quad .$$

Thus

$$B(b,c_o) = \frac{2b^2}{n} + \bar{\kappa}\,\bar{b}^2 + \kappa\bar{b}^2 \left[\frac{2A^2}{A^2 + 1} - 1 \right]^2$$

$$= \frac{2b^2}{n} + \frac{\bar{b}^2}{n' + 2}\,(2 + n'G^2) \tag{27}$$

where

$$G \equiv \frac{A^2 - 1}{A^2 + 1} \; .$$

Differentiating to minimize $B(b,c_o)$ over values of b, the minimizing b is

$$b_o = \frac{\dfrac{2 + n'G^2}{n' + 2}}{\dfrac{2}{n} + \dfrac{2 + n'G^2}{n' + 2}} \; , \tag{28}$$

and the Λ-minimax rule is

$$\delta_o(w) = b_o w + \bar{b}_o 2 \left[\frac{A}{A^2 + 1} \right] \left[\frac{n'}{n' + 2} \right] \Delta$$

which is the Bayes rule

$$\delta(w) = \frac{\dfrac{w}{n' + 2} + \dfrac{v'}{n}\left(\dfrac{n'}{n' + 2} \right)}{\dfrac{1}{n' + 2} + \dfrac{1}{n}}$$

when $A = 1$. (The posterior distribution of $\tilde{h}$ is gamma-2 with parameters

$$n'' = n' + n \quad \text{and} \quad v'' = \frac{nw + n'v'}{n' + n} = \frac{\dfrac{w}{n'} + \dfrac{v'}{n}}{\dfrac{1}{n'} + \dfrac{1}{n}} \quad .$$

The posterior mean is $\frac{1}{v''}$, and so the Bayes estimate of $\tilde{\sigma}^2$ for quadratic loss is v''. Recall that the loss function for this problem, $h^2(\delta(w) - \frac{1}{h})^2$ is not quadratic, which accounts for the apparent discrepancy.)

To determine the efficiency of the Λ-minimax rule with respect to the Bayes rule, compute the risk of the Λ-minimax rule by substituting b_o of (28) for b in $B(b,c_o)$ of (27). The result is

$$B(b_o,c_o) = \frac{2(2 + n'G^2)}{2(n' + 2) + n(2 + n'G^2)} \quad . \tag{29}$$

Note that this does not depend on Δ. The Bayes risk (of the Bayes rule) is obtained by substituting A = 1 (and therefore G = 0) in $B(b_o,c_o)$ giving

$$\frac{2}{n + n' + 2} \; .$$

The efficiency of the Λ-minimax rule with respect to the Bayes rule is taken to be the ratio of these two risks. So

$$\text{Eff.} = \frac{\dfrac{2}{n + n' + 2}}{\dfrac{2(2 + n'G^2)}{2(n' + 2) + n(2 + n'G^2)}}$$

$$= \frac{2(n' + 2) + n(2 + n'G^2)}{(n + n' + 2)(2 + n'G^2)} \; .$$

Figures 1-4 give contours of the efficiency for sample sizes n = 2, 5, 10, 20. For given n, the label of the contour at the point (A,n') is the efficiency of the Λ-minimax rule with respect to the Bayes rule when the degrees of freedom parameter of the prior distribution is n', and the incompleteness specification U, is determined by A. Only for large values of n', do increases in A significantly decrease the efficiency, and the importance of small values of A decreases as the sample size increases.

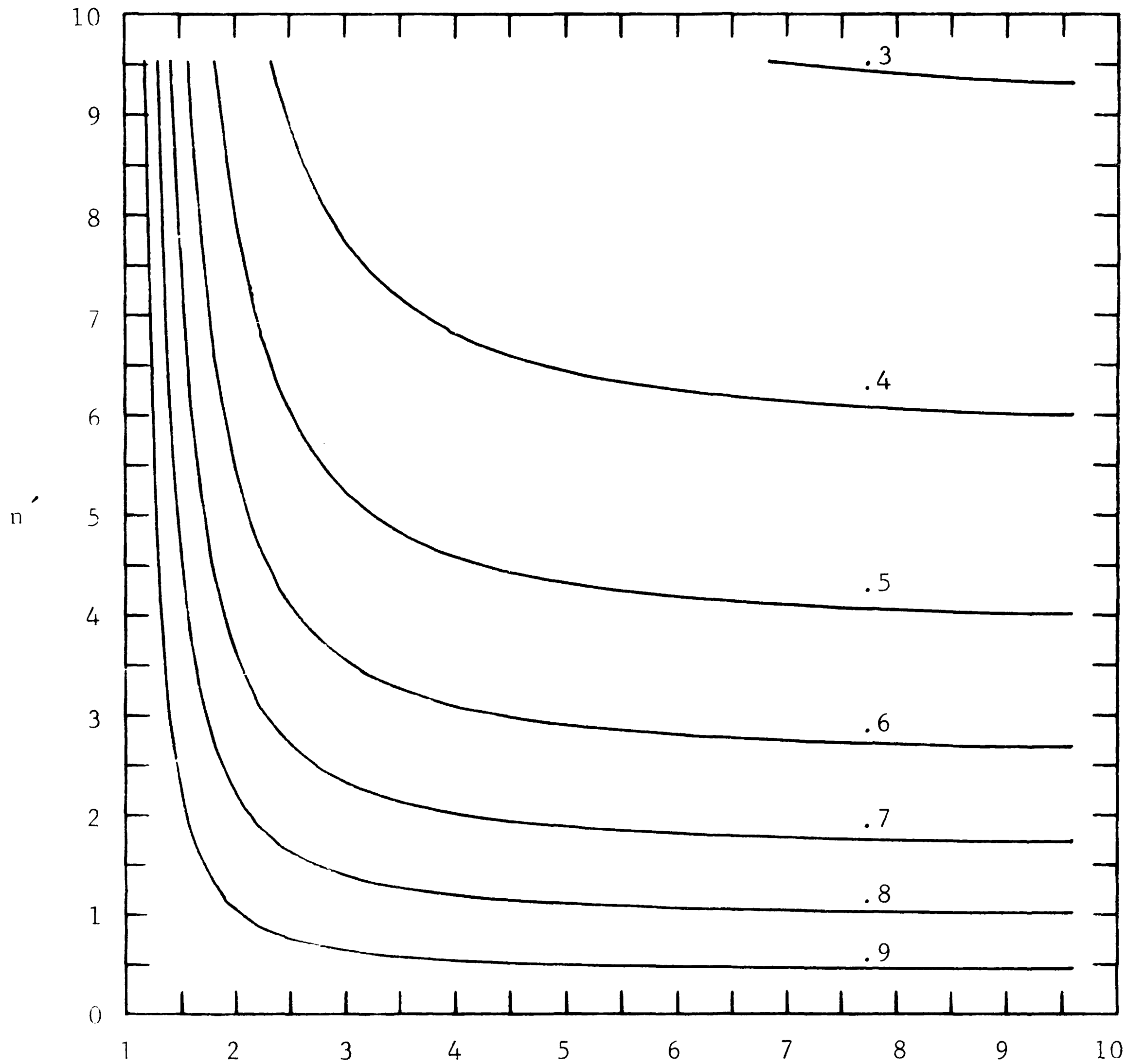

Fig. 1.--Contours of: (risk of Bayes estimate)/(risk of Λ-minimax estimate) for estimation of a scale parameter. The degrees of freedom in the prior distribution is n', the incompleteness specification is determined by A, and the sample size is $n = 2$.

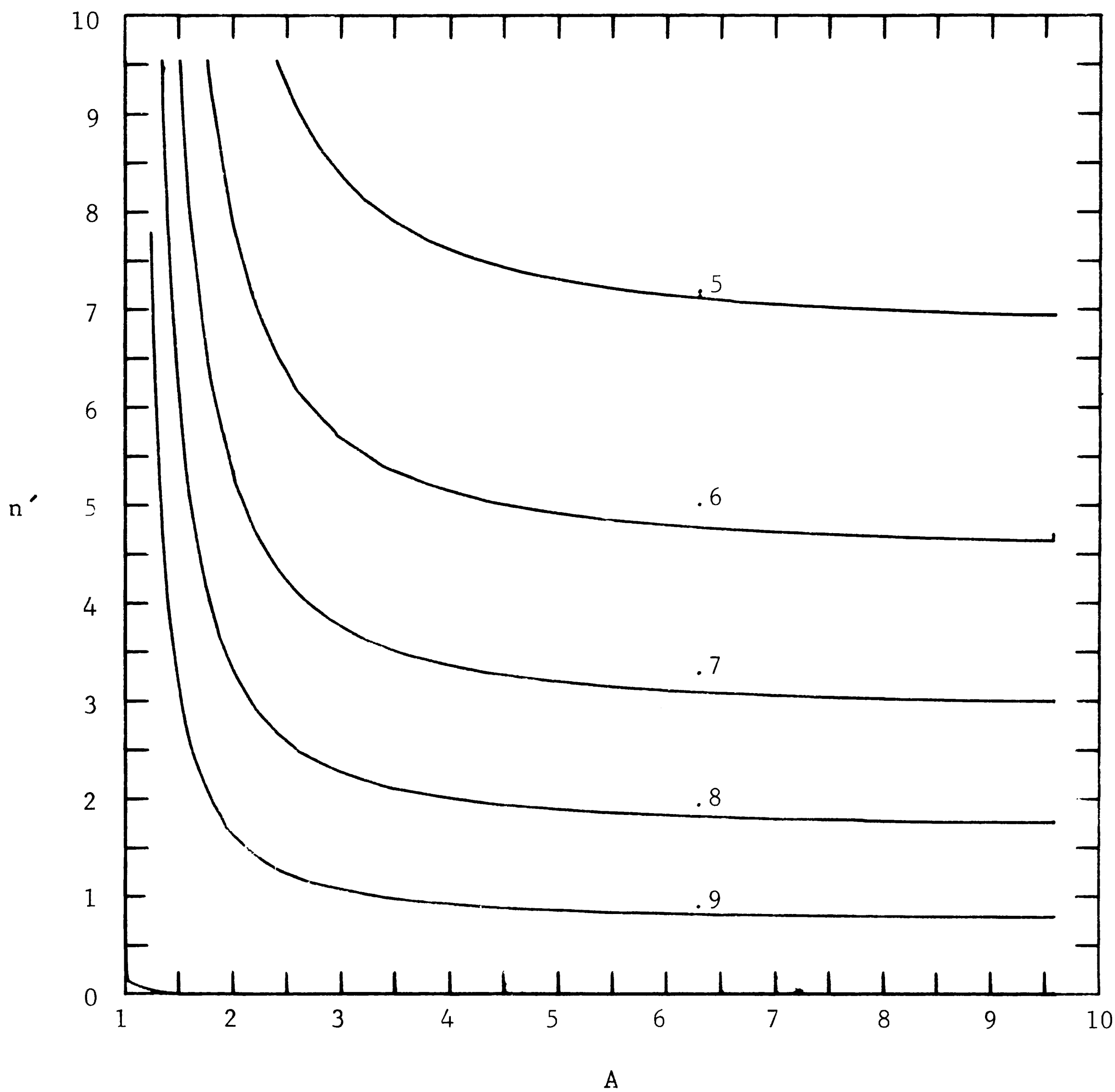

Fig. 2.--Contours of: (risk of Bayes estimate)/(risk of Λ-minimax estimate) for estimation of a scale parameter. The degrees of freedom in the prior distribution is n′, the incompleteness specification is determined by A, and the sample size is n = 5.

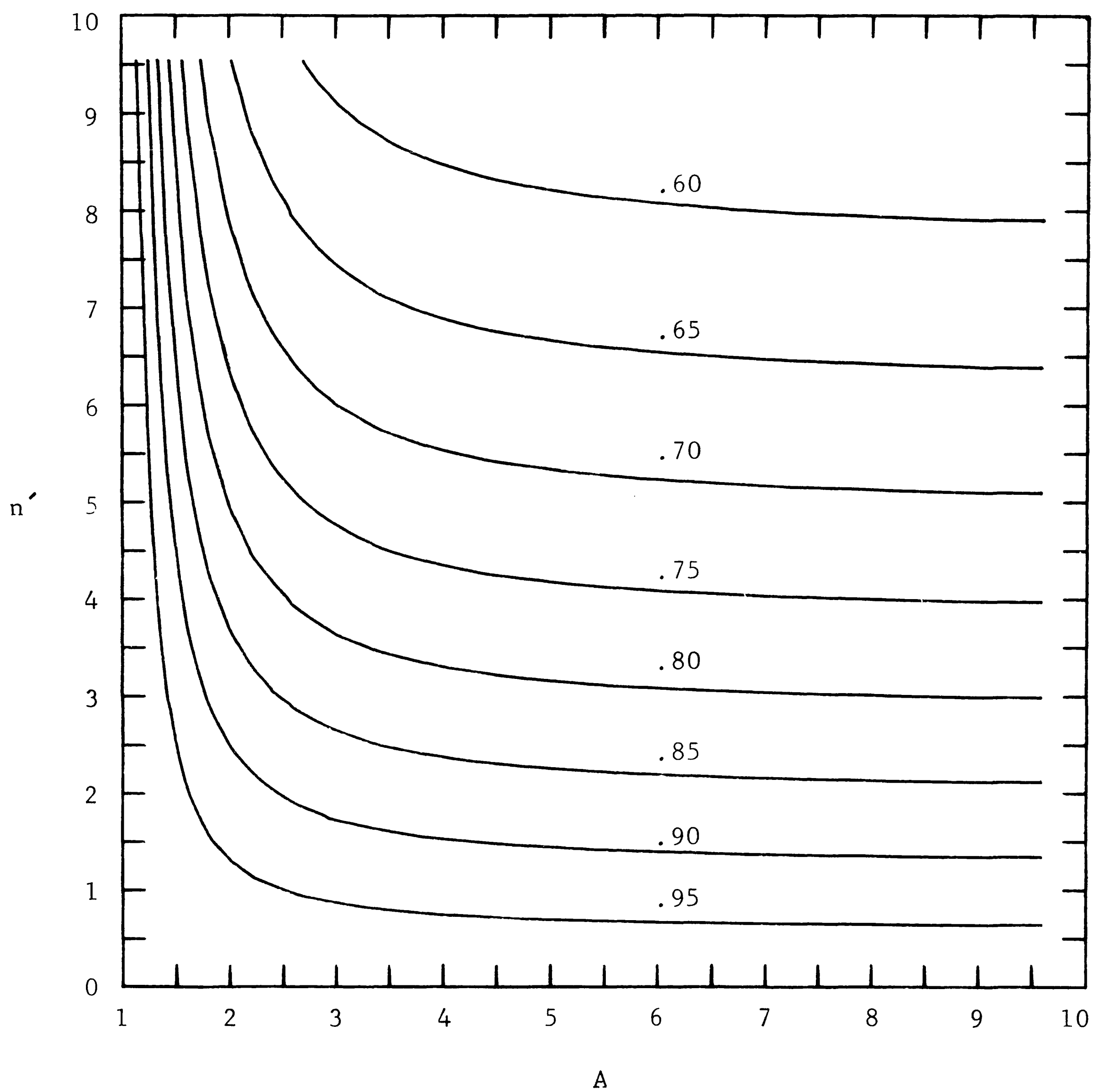

Fig. 3.--Contours of: (risk of Bayes estimate)/(risk of Λ-minimax estimate) for estimation of a scale parameter. The degrees of freedom in the prior distribution is n′, the incompleteness specification is determined by A, and the sample size is n = 10.

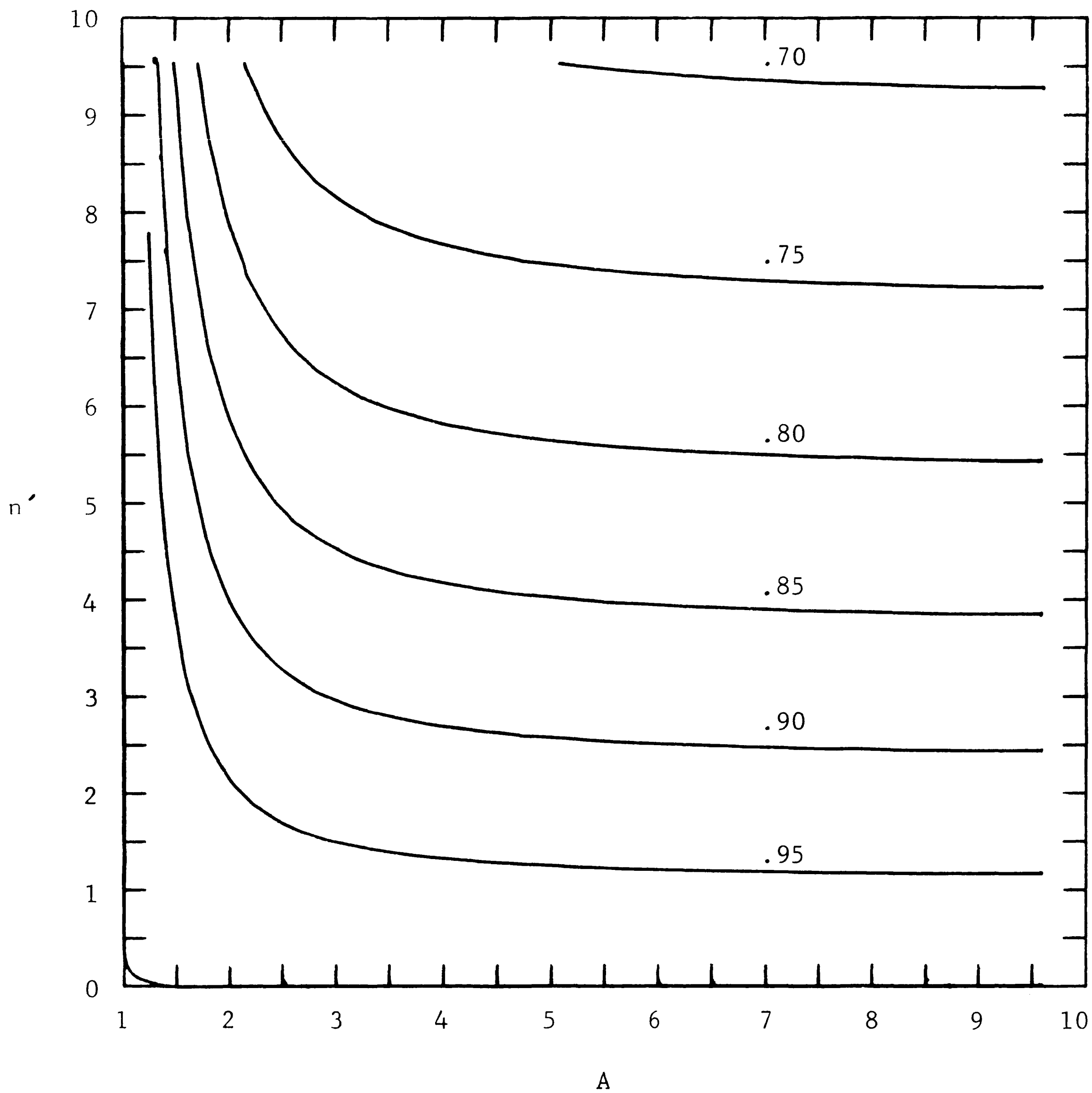

Fig. 4.--Contours of: (risk of Bayes estimate)/(risk of Λ-minimax estimate) for estimation of a scale parameter. The degrees of freedom in the prior distribution is n′, the incompleteness specification is determined by A, and the sample size is n = 20.

The graphs were obtained by evaluating the efficiency over a 2,250 point grid at $A = 1(.2)9.8$ and $n' = 0(.2)9.8$ on a CDC6400 computer, and having the contours smoothed and plotted on an EAI3500 Dataplotter.

The results of this section extend immediately to the mean non-zero. In section 5.2 the estimation of $\tilde{\sigma}^2$ when $\underset{\sim}{\bar{\theta}}$ is not known is treated.

5.2 Mean Unknown

Recall the problem of section 4.3 in which $X_1, X_2, \cdots, X_n$ are independent, identically Normally distributed random variables with mean θ and variance h^{-1}. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $V = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, where $V = 0$ if $n \le 1$. The joint density of $\bar{X}$ and V is

$$f(\bar{x},v|\theta,h) = (\text{const.})\left[e^{-\frac{1}{2}hn(\bar{x}-\theta)^2}h^{\frac{1}{2}}\right]\left[e^{-\frac{1}{2}hrv}h^{\frac{1}{2}r}\right]$$

where $r = n - 1$, and the natural conjugate prior for $\underset{\sim}{\bar{\theta}}$ and $\tilde{h}$ is the Normal-gamma with density

$$\begin{aligned} f_{N_\gamma}(\theta,h|\mu,v',n^*,n') &= f_N(\theta|\mu,hn^*)f_{\gamma_2}(h|v',n') \\ &= (\text{const.})\left[e^{-\frac{1}{2}hn^*(\theta-\mu)^2}h^{\frac{1}{2}}\right]\left[e^{-\frac{1}{2}hn'v'}h^{\frac{1}{2}n'-1}\right] \end{aligned} \tag{30}$$

for $-\infty \le \theta \le \infty$, $h \ge 0$, and $v',n^*,n' > 0$.

The following moments will be required:

$$\begin{aligned} E(V|\tilde{h}) &= \tilde{h}^{-1}, & \text{var}(V|\tilde{h}) &= \frac{2}{r\tilde{h}^2}, \\ E(\tilde{h}) &= \frac{1}{v'}, & \text{var}(\tilde{h}) &= \frac{2}{n'v'^2}. \end{aligned} \tag{31}$$

Suppose that h^{-1} is to be estimated by a function δ of the sufficient statistic, V, with loss

$$h^2(\delta(v) - h^{-1})^2$$

as in the previous section. Assume that prior knowledge about v' is incomplete but that the decision maker can specify

$$v' \in U = \{v' \mid \frac{\Delta}{A} \le v' \le \Delta A\} ,$$

where $\Delta \ge 0$, $A \ge 1$. Here Λ is the set of Normal-gamma distributions (30), with $v' \in U$.

Restricting the discussion to rules linear in V, write

$$\delta(v) = bv + \bar{b}c \frac{n'}{n' + 2} \Delta$$

as in the previous section and determine b,c to minimize

$$\sup_{v' \in U} E\, \tilde{h}^2(\delta(V) - \tilde{h}^{-1})^2 = \sup_{v' \in U} E\, \tilde{h}^2 \left[bV + bc \frac{n'}{n' + 2} \Delta - \tilde{h}^{-1}\right]^2 .$$

The expectations in display (31) are the same as those in (24) with W replaced by V, and n in $\mathrm{var}(W|\tilde{h}) = \frac{2}{n\tilde{h}^2}$ by $r = n - 1$ in $\mathrm{var}(V|\tilde{h}) = \frac{2}{r\tilde{h}^2}$. Thus with these substitutions, the required computations are identical to those of section 5.1, and hence the Λ-minimax rule when $\underline{\bar{\theta}}$ is unknown is

$$\delta(v) = b_1 v + \bar{b}_1 2 \frac{A}{A^2 + 1} \frac{n'}{n' + 2} \Delta ,$$

where

$$b_1 = \frac{\dfrac{2 + n'G^2}{n' + 2}}{\dfrac{2}{r} + \dfrac{2 + n'G^2}{n' + 2}} ,$$

and

$$G = \frac{A^2 - 1}{A^2 + 1} .$$

Furthermore, the Λ-minimax risk of the Λ-minimax rule is, by comparison with (29),

$$\frac{2(2 + n'G^2)}{2(n' + 2) + r(2 + n'G^2)} . \tag{32}$$

It follows that the contours given in figures 1 - 4, of the efficiency of the Λ-minimax rule with respect to the Bayes rule, can be used when θ is unknown if n is replaced by $r = n - 1$.

6. ACKNOWLEDGMENTS

The author wishes to acknowledge Mr. Joseph Cameron for a helpful discussion and Professor I. Richard Savage for considerable advice on this research. Financial support was made available by several sources: The National Science Foundation through a grant to the Florida State University Computing Center, The National Institutes of Health through its Biometry Training Grant program, and The Office of Naval Research.

REFERENCES

[1] Anderson, T. W. and Gupta, S. Das. "Some inequalities on characteristic roots of matrices," Biometrika, 50 (1963), 522-524.

[2] Blum, J. R. and Rosenblatt, J. "On partial a priori information in statistical inference," Annals of Mathematical Statistics, 38 (1967), 1671-1678.

[3] George, S. L. "Partial prior information: some empirical Bayes and G-minimax decision functions," THEMIS Technical Report 48, Department of Statistics, Southern Methodist University, (1969).

[4] Hadley, G. Nonlinear and Dynamic Programming, Reading, Mass.: Addison-Wesley, (1964).

[5] Hamer, W. J. Standard Cells, National Bureau of Standards Monograph 84, (1965).

[6] Hodges, J. L., Jr. and Lehmann, E. L. "The use of previous experience in reaching statistical decisions," Annals of Mathematical Statistics, 23 (1952), 396-407.

[7] Karlin, S. Mathematical Methods and Theory in Games, Programming, and Economics, Vol. 2, Reading, Mass.: Addison-Wesley, (1959).

[8] Menges, G. "On the 'Bayesification' of the minimax principle," Unternehmensforschung, 10, No. 2, (1966), 81-91.

[9] Raiffa, H. and Schlaifer, R. Applied Statistical Decision Theory, Division of Research, Graduate School of Business Administration, Harvard University, Boston, (1961).

[10] Randles, R. H. "Γ-minimax procedures for the use of incomplete prior information in selection and classification problems," Doctoral Dissertation, Florida State University, (1969).

[11] Rao, C. R. Linear Statistical Inference and Its Applications, New York: John Wiley and Sons, Inc., (1965).

[12] Schneeweiss, H. "Eine Entscheidungsregel für den Fall partiell bekannter Wahrscheinlichkeiten," Unternehmensforschung, 8, No. 2 (1964), 86-95.

[13] Skibinsky, M. and Cote, L. "On the inadmissibility of some standard estimates in the presence of prior information," Annals of Mathematical Statistics, 34 (1963), 539-548.

[14] Stein, C. "Statistical decision theory," Unpublished notes, Stanford University, (1963).

[15] Stone, M. "Robustness of non-ideal decision procedures," Journal of the American Statistical Association, 58 (1963), 480-486.

[16] Thompson, J. R. "Some shrinkage techniques for estimating the mean," Journal of the American Statistical Association, 63 (1968), 113-122.

[17] Winkler, R. "The assessment of prior distributions in Bayesian analysis," Journal of the American Statistical Association, 62 (1967), 776-800.

APPENDIX OF MATHEMATICAL PROOFS

Lemma 3: The set β (11) is convex, and for each $e \in \mathcal{E}$ (12), $R(e,B)$ (8) is convex in B.

Proof: Given B_o, $B_1 \in \beta$ and $0 < \alpha < 1$, let $B_\alpha = (1 - \alpha)B_o + \alpha B_1$. It must be shown that for each $e \in \mathcal{E}$

$$R(e,B_\alpha) \leq (1 - \alpha)R(e,B_o) + \alpha R(e,B_1) .$$

It follows from (8) and properties of the trace that

$$R(e,B) = \text{tr } B^T K B \Sigma_1 + \text{tr } \bar{B}^T K \bar{B}(\Sigma_0 + ee^T) .$$

We wish to show that $T \leq 0$ where

$$T = R(e,B_\alpha) - (1 - \alpha)R(e,B_o) - \alpha R(e,B_1) .$$

Let $\Sigma = \Sigma_0 + ee^T$. Then it can be shown that

$$\begin{aligned} T &= \text{tr}[B_\alpha^T K B_\alpha \Sigma_1 + \bar{B}_\alpha^T K \bar{B}_\alpha \Sigma] - (1 - \alpha)\text{tr}[B_o^T K B_o \Sigma_1 + \bar{B}_o^T K \bar{B}_o \Sigma] - \alpha\text{tr}[B_1^T K B_1 \Sigma_1 + \bar{B}_1^T K \bar{B}_1 \Sigma] \\ &= -\alpha(1 - \alpha)\text{tr}(B_o - B_1)^T K(B_o - B_1)(\Sigma_1 + \Sigma) . \end{aligned}$$

It will suffice to show that

$$S = \text{tr}\{[(B_o - B_1)^T K(B_o - B_1)][\Sigma_1 + \Sigma]\} \geq 0 .$$

Now $\Sigma_1 + \Sigma = \Sigma_1 + \Sigma_0 + ee^T$ is symmetric positive definite, $(B_o - B_1)^T K(B_o - B_1)$ is symmetric non-negative definite, and thus all the characteristic roots of each are non-negative. See Rao [11, p. 35]. It follows from a theorem of Anderson and Gupta [1, corollary 2.2.1] that all the characteristic roots of their product are non-negative. Since the trace of a matrix is the sum of its characteristic roots, $S \geq 0$. Hence for all $e \in \mathcal{E}$

$$R(e,B_\alpha) \leq (1 - \alpha)R(e,B_o) + \alpha R(e,B_1) . \tag{33}$$

It remains to show that $B_\alpha \in \beta$, that is β is convex. By (13)

$$\beta = \{B \in \mathcal{T}_p \mid \sup_{\mu \in U} R(\mu,B) \le \tau\} = \{B \in \mathcal{T}_p \mid \sup_{e \in \mathcal{E}} R(e,B) \le \tau\} .$$

By (33) with B_o, $B_1 \in \beta$, it follows that

$$\begin{aligned} \sup_{e\in\mathcal{E}} R(e,B_\alpha) &\le \sup_{e\in\mathcal{E}}[(1-\alpha)R(e,B_o) + \alpha R(e,B_1)] \\ &\le (1-\alpha)\sup_{e\in\mathcal{E}} R(e,B_o) + \alpha \sup_{e\in\mathcal{E}} R(e,B_1) \\ &\le (1-\alpha)\tau + \alpha\tau = \tau . \end{aligned}$$

That is, $B_\alpha \in \beta$. Thus, the lemma.

<u>Lemma 6</u>: The U-minimax risk (in L) satisfies

$$\inf_{B\in\mathcal{T}_p} \sup_{\mu\in U} R(\mu,B) = \sup_{\Pi\in\tilde{\mathcal{E}}} \inf_{B\in\mathcal{T}_p} R(\Pi,B) .$$

<u>Proof</u>:

$$\begin{aligned} \inf_{B\in\mathcal{T}_p} \sup_{\mu\in U} R(\mu,B) &= \inf_{B\in\beta} \sup_{\mu\in U} R(\mu,B) && \text{by corollary 2.1} \\ &= \inf_{B\in\beta} \sup_{\Pi\in\tilde{\mathcal{E}}} R(\Pi,B) && \text{by lemma 5} \\ &= \sup_{\Pi\in\tilde{\mathcal{E}}} \inf_{B\in\beta} R(\Pi,B) && \text{by theorem 4} \qquad (34) \end{aligned}$$

Suppose there exists a $B^* \in \mathcal{T}_p$ such that

$$\sup_{\Pi\in\tilde{\mathcal{E}}} \inf_{B\in\mathcal{T}_p} R(\Pi,B) = \sup_{\Pi\in\tilde{\mathcal{E}}} R(\Pi,B^*) . \qquad (35)$$

By theorem 4, there exists an $\tilde{\mathcal{E}}$-minimax rule $B_o \in \beta$ so

$$\sup_{\mu \epsilon U} R(\mu, B^*) = \sup_{\Pi \epsilon \tilde{\mathcal{E}}} R(\Pi, B^*) \qquad \text{by lemma 5}$$

$$= \sup_{\Pi \epsilon \tilde{\mathcal{E}}} \inf_{B \epsilon \mathcal{T}_p} R(\Pi, B) \qquad \text{by (35)}$$

$$\leq \sup_{\Pi \epsilon \tilde{\mathcal{E}}} \inf_{B \epsilon \beta} R(\Pi, B) \qquad \text{since } \beta \subset \mathcal{T}_p$$

$$= \inf_{B \epsilon \beta} \sup_{\Pi \epsilon \tilde{\mathcal{E}}} R(\Pi, B) \qquad \text{by theorem 4}$$

$$= \sup_{\Pi \epsilon \tilde{\mathcal{E}}} R(\Pi, B_o) \qquad \text{since } B_o \text{ is } \tilde{\mathcal{E}}\text{-minimax}$$

$$\leq \tau \qquad \text{since } B_o \epsilon \beta .$$

Thus $B^* \epsilon \beta$. It has been shown that if $B^* \epsilon \mathcal{T}_p$ and B^* satisfies (35), then $B^* \epsilon \beta$. Thus

$$\sup_{\Pi \epsilon \tilde{\mathcal{E}}} \inf_{B \epsilon \beta} R(\Pi, B) = \sup_{\Pi \epsilon \tilde{\mathcal{E}}} \inf_{B \epsilon \mathcal{T}_p} R(\Pi, B) ,$$

and the result follows from (34).

<u>Lemma 8</u>: For p-vectors $G \neq 0$, μ, and $M \geq 0$, $p \times p$ matrix $\bar{B}$ and positive definite $p \times p$ matrix K,

$$\operatorname*{supremum}_{|\mu| \leq M} [\bar{B}\mu + G]^T K [\bar{B}\mu + G] > \operatorname*{supremum}_{|\mu| \leq M} [\bar{B}\mu]^T K [\bar{B}\mu] . \tag{18}$$

<u>Proof</u>: There is a μ_o such that $|\mu_o| \leq M$ and the supremum on the right-hand side of (18) is attained at μ_o. Now

$$[\bar{B}\mu_o + G]^T K [\bar{B}\mu_o + G] = \mu_o^T \bar{B}^T K \bar{B} \mu_o + 2\mu_o^T \bar{B}^T K G + G^T K G ,$$

and since $|-\mu_o| \leq M$

$$\operatorname*{supremum}_{|\mu| \leq M} [\bar{B}\mu + G]^T K [\bar{B}\mu + G] \geq \mu_o^T \bar{B}^T K \bar{B} \mu_o + 2|\mu_o^T \bar{B}^T K G| + G^T K G .$$

But K is positive definite and $G \neq 0$, so $G^T K G > 0$ and thus

$$\underset{|\mu| \le M}{\text{supremum}}[\bar{B}\mu + G]^T K[\bar{B}\mu + G] > \mu_o^T \bar{B}^T K \bar{B} \mu_o = \underset{|\mu| \le M}{\text{supremum}}[\bar{B}\mu]^T K[\bar{B}\mu] \ .$$

<u>Lemma 11</u>: With B_o as in (20), there exists a distribution λ_o on $\mathcal{E}$ such that

$$R = E_{\lambda_o} R(e,B_o) = \inf_{B \in \mathcal{T}_p} E_{\lambda_o} R(e,B) \ .$$

<u>Proof</u>: For $e_i \in \mathcal{E}$ let

$$\lambda_o(e_i) = \frac{1}{u} \ , \quad i = 1, 2, \cdots, u \ .$$

Then for $B \in \mathcal{T}_p$

$$E_{\lambda_o} R(e,B) = \sum_{i=1}^{u} R(e_i,B)\lambda_o(e_i)$$

$$= \operatorname{tr} B^T K B \Sigma_1 + \operatorname{tr} \bar{B}^T K \bar{B} \Sigma_0 + \frac{1}{u} \sum_{i=1}^{u} e_i^T \bar{B}^T K \bar{B} e_i$$

$$= \operatorname{tr} B^T K B \Sigma_1 + \operatorname{tr} \bar{B}^T K \bar{B} \Sigma_0 + \frac{1}{u} \operatorname{tr} \bar{B}^T K \bar{B} \sum_{i=1}^{u} e_i e_i^T \ .$$

With some computation it can be shown that

$$\sum_{i=1}^{u} e_i e_i^T = u\hat{M}^2 \ ,$$

where $\hat{M}$ is the $p \times p$ diagonal matrix with entries $m_1, m_2, \cdots, m_p$. Thus

$$E_{\lambda_o} R(e,B) = \operatorname{tr} B^T K B \Sigma_1 + \operatorname{tr} \bar{B}^T K \bar{B}(\Sigma_0 + \hat{M}^2) \ .$$

The calculus may be applied directly to determine the minimizing B, but note that the minimization is identical to that in the proof of theorem 7 if D_Π is replaced by $\hat{M}^2$ in display (16). Thus from (17), the minimizing B is

$$(\Sigma_0 + \hat{M}^2)(\Sigma_1 + \Sigma_0 + \hat{M}^2)^{-1} = \text{diag.} \left[\frac{\sigma_i^0 + m_i^2}{\sigma_i^1 + \sigma_1^0 + m_i^2} \right] = B_0 \;.$$